DEPARTMENT OF DEFENSE

DEFENSE
LOGISTICS
AGENCY

Cameron Station,
Alexandria, Virginia 22304-6100

A STUDY OF DEMAND FORECASTING IN THE DEFENSE LOGISTICS AGENCY

Operations Research and Economic Analysis Office

FEBRUARY 1986

A STUDY OF
DEMAND FORECASTING IN THE
DEFENSE LOGISTICS AGENCY

February 1986

Dr. Stan Orchowsky, Capt Ronald Kirchoff,
Ms. Janet Rider, Cpt(P) Dale Kem
Operations Research and Economic Analysis Office
Headquarters, Defense Logistics Agency
Cameron Station, Alexandria, Virginia

DEFENSE LOGISTICS AGENCY
HEADQUARTERS
CAMERON STATION
ALEXANDRIA, VIRGINIA 22314

DLA-LO

Feb 1986

FOREWORD

This report presents the findings of the study of demand forecasting conducted by the Operations Research and Economic Analysis Office at DLA. The study compared a number of different forecasting methods to determine if improvements over the current DLA forecasting method could be obtained. The methods were compared using both forecast error and impacts on inventory system variables as criteria for judging improvement.

The results of the study showed that the preferred method produced a 3.9% decrease in the average forecast error over the current system. Positive impacts on safety level dollars and other inventory variables would also be realized if, as the study recommends, this alternative technique is implemented. The report also offers several other recommendations for improving demand forecasting in DLA.

ROGER C. ROY
Acting Assistant Director,
Policy and Plans

CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This report documents the findings of the study of demand forecasting conducted by DLA's Operations Research and Economic Analysis Office. The goal of the study was to identify alternative methods which would increase the accuracy of DLA's demand forecasts. The initial phase of the study was a literature review of a wide range of potential forecasting techniques to determine their applicability to DLA's forecasting needs. Based on this review, 17 forecasting techniques were identified which showed promise in being useful alternatives to the current method used by DLA.

The next step in the analysis was to compare the accuracy of the 18 forecasting methods (the current DLA method plus the 17 alternative methods) using a random sample of 6,412 items. A maximum of eight years of historic demand data was available for these items. The result of these preliminary analyses was the identification of six methods which appeared to be the best performers.

The analysis then examined the forecast accuracy of these six methods, both individually and in combinations. Two approaches for combining methods were examined. The first used unweighted and weighted averages of the forecasts produced by the different methods. The other procedure involved the use of item characteristics to predict which of a group of forecast methods would be most accurate for each item. The results showed that the best average consisted of the forecasts from single exponential smoothing and a four-quarter moving average. The prediction of item groupings was not as successful, but the best of these methods was retained for further analysis.

The above results were validated with two additional samples of items. The results showed that a weighted average of the forecasts of single exponential smoothing and the four-quarter moving average produced the best results across the three samples. Several of these methods were then tested on the entire population of 636,000 items. The results showed that the weighted average produced a 3.9% decrease in the average forecast error when compared with the exponential smoothing method currently used in SAMMS.

A simulation analysis was then conducted in order to obtain some preliminary data regarding the performance in SAMMS of methods which were statistically better than the current SAMMS method. The simulation examined the impacts of these methods on five inventory variables: supply availability, safety level dollars, total dollar commitments, number of backorders, and number of days on backorder. The results confirmed the superior performance of the weighted average model, which consistently produced positive impacts on these inventory variables. However, the results of the simulation analysis revealed several key issues regarding how the new forecasting procedure should be implemented. Further study would be required to examine these issues and determine how best to implement a new method so as to obtain the maximum benefit as quickly as possible.

The study concludes that the weighted average of the single exponential smoothing and the four-quarter moving average forecasts is the best of the forecasting methods examined for all commodities except Medical. For Medical, the four-quarter moving average alone is the best method. Based on a decrease in safety level dollars proportional to a decrease in forecast error, improvement of the best method over the current method is estimated to be as follows:

Commodity	Percent Reduced Error	Estimated Reduced Safety Level ($)
Construction	1.1%	$ 2,715,950
Electronics	1.6	1,912,796
General	6.0	12,714,489
Industrial	4.7	5,885,924
Medical	3.7	607,486
C & T	1.4	1,586,742

It is recommended that this method be implemented in SAMMS, following additional study regarding how to best incorporate the model into the system.

One additional recommendation was made based on supplemental analyses documented in the report. All items, including VIP items, should be forecasted on a quarterly basis. This change would result in large decreases in safety level dollars and total commitments with no noticeable change in supply availability.

I. INTRODUCTION

A. BACKGROUND

DLA currently uses in its Standard Automated Material Management System (SAMMS) a single model to forecast demand for all items, with the exception of Program Oriented Items (POI) and Government Furnished Materiel (GFM). DLA-LO's 1981 Backorder Review found that DLA's inability to forecast demand changes was the primary cause of backorders. Since this finding, several directed actions on demand forecasting have occurred, both at the Headquarters and at the Primary Level Field Activities. There is a general consensus that improved forecasting could result in improved Agency mission performance and reduced costs. Based on the findings of a recent subsistence demand forecasting study, it is felt that potential increases in forecast accuracy may be obtained by applying alternative methods of forecasting techniques to different categories of items.

B. PROJECT DEFINITION

1. Statement of Problem

Currently, DLA uses a single model to forecast demand for all hardware items. The potential exists for improved forecasts using new and different forecasting models for different categories of items.

2. Purpose of Project

The purpose of this project was to study various techniques for forecasting demand of all DLA commodities, except subsistence and fuels, and to determine whether forecast accuracy can be improved by applying alternative forecasting techniques to different categories of items.

3. Specific Objectives. The specific objectives of this study are:

(a) To evaluate both classical and innovative forecasting techniques as to their applicability to forecasting DLA's demand.

(b) To determine whether different techniques applied to different categories of DLA's items would produce lower forecast error than the single method currently used by SAMMS.

(c) To examine the effects of applicable forecasting techniques on DLA's inventory management system.

(d) To provide recommendations for improvement of DLA's current forecasting procedures, and for implementation of alternative methods if appropriate.

C. SCOPE OF PROJECT

1. Project Effort

(a) All DLA commodities, except subsistence and fuels, were included in the study.

(b) All stocked items were examined except for new items, POI, and GFM. Items which were classified as Numeric Stockage Objective (NSO) for a significant portion of their time in the system were also excluded.

2. Report Organization

This study was conducted in three phases, over the course of one year. At the end of each phase, an interim report was prepared which documented the results of the analyses to that point. The current report represents a synthesis of the contents of the three interim reports.

The current report is divided into eight major sections, including this first introductory section. Section II presents a review of the literature, which includes discussions of forecasting in DLA, in the Services, and in the academic literature.

Section III presents the results of the review of forecasting techniques to be included in the study. Brief discussions of each technique and its merits are presented in this section.

Section IV describes the methodology and procedures of the study. Included here is a description of the data used, the selection of item samples, and the procedures used to implement the forecast methods.

Section V presents the findings of the data analysis. This includes the results of preliminary analyses designed to identify a single number of potentially useful methods from the larger group of procedures identified in Section III, the assessment of these procedures with regard to their accuracy, the validation of these results using additional samples and the entire population, and an assessment of the impacts of these alternative methods on the inventory system.

The next section presents a summary of the findings and a discussion of the results. Sections VII and VIII present the conclusions and recommendations (respectively) resulting from the data analysis.

II. LITERATURE REVIEW

A. Current DLA Forecasting Method

The current DLA forecasting method is described in DLAM 4140.2, Vol. II, Chapter 53, "Recurring Demand Forecast." This chapter describes the forecast computations and the items to which this method is applied.

The method currently used by DLA is a version of Brown's double exponential smoothing. The smoothing is carried out by depot location, but since this is not crucial to the present study, the locations will not be discussed here. The formulas used by DLA are as follows:

$$S'_t = \alpha X_t + (1-\alpha)S'_{t-1}$$

$$S''_t = \alpha S'_t + (1-\alpha)S''_{t-1}$$

$$a_t = 2S'_t - S''_t \quad ,$$

where X_t is the demand for an item for time period t, S'_t is the single exponential smoothed value for the current time period t, S''_t is the double exponential smoothed value for time period t, α is the smoothing constant, and a_t is the expected value of the demand data at time t.

Exponential smoothing thus weighs the current actual demand value and the previous smoothed demand to develop the expected demand for the next time period. Alpha is the weight used in this process, and is normally .2 for most DLA items.

One aspect of the formulas presented above deserves discussion. The value a_t is used by DLA as the forecast for the next time period. In the original formulation of double exponential smoothing, however, a_t is not intended to be the forecast value. R. G. Brown, generally acknowledged as the developer of exponential smoothing, makes this clear in his presentation of double exponential smoothing (1, pp. 128-132). The value a_t is merely the estimate of the current level of the demand series. To this must be added an estimate of the trend, b_t

$$b_t = (\alpha/1-\alpha)\ (S'_t - S''_t)$$

The forecast is then given by

$$F_{t+m} = a_t + b_t m,$$

where m is the number of periods ahead to be forecast.

This misconception was addressed recently in an article by Gardner (2), who notes that the use of the a_t term as the forecast is a common mistake in the literature. One advantage of double exponential smoothing is that it is appropriate for data in which a trend exists. The result of using the a_t term as the forecast, however, is that forecasts will consistently lag any trend in the data (2). Thus one point of interest in the current study will be to compare the accuracy of DLA's current method with the double exponential smoothing method as originally proposed. This is discussed more fully in Section III of the report.

B. Previous DLA Studies

Several studies exist which have addressed the issue of forecasting in DLA. These will be reviewed briefly in this section.

The original study which recommended using exponential smoothing as the standard DLA (then DSA) forecasting method was conducted in 1963 (3). Part of the study used simulated monthly demand to compare five different forecasts: a 4-quarter weighted moving average, single exponential smoothing (with trend correction) using alpha values of .2 and .4, and double exponential smoothing using alpha values of .2 and .4 (it should be noted that in this study both the single and double exponential smoothing formulas failed to take into account the trend term; this is the same problem that was discussed previously). The performance measures examined were the average investment per item, and the percentage of demand filled without backorders. The results of the comparison showed that both exponential smoothing methods were superior to the moving average, although there was no difference between the two smoothing methods. Using the alpha value of .2 produced more accurate forecasts than using a value of .4.

A second study was conducted several years later to verify the findings of the original study (4). This second study used a combination of actual and simulated data to examine three different demand patterns: level, trend (with varying slopes) and modified trend. Four forecasting methods were compared: single exponential smoothing with a tracking signal, double exponential smoothing, a 4-quarter weighted moving average, and a 4-quarter unweighted moving average. The performance criteria included the mean absolute deviation (MAD) of the forecast errors, and the mean percentage error (MPE) of the forecasts. The results of this analysis showed that double exponential smoothing with an alpha value between .1 and .2 produced the most accurate forecasts. This study, along with the earlier one, seem to have established the forecasting method still used by DLA.

Over the years, several forecasting studies have been done by students using data from one or more commodities. Typical of these studies is one done by Praggy (5). The study used 12 quarters of data from the electronics commodity and compared a variety of forecasting methods, including single and double exponential smoothing, various weighted and unweighted moving averages, using the mean of the data and using the last observation as forecasts (known as "naive" methods), and polynomial fitting. The results showed that for replenishment items, the best forecasts were produced by simple exponential smoothing with alphas between .2 and .4. High demand items (200 or more quarterly demands) were best forecasted as using a 4-quarter moving average, while medium and low demand items (20-200 demands and under 20 demands, respectively) were best forecasted using single exponential smoothing with alphas between .1 and .4.

Several DLA forecasting studies have either been recently completed or are currently in progress. One of these is being conducted at the Defense Electronics Supply Center (DESC; 6). The interim report presented findings which examined various versions of exponential smoothing for 36 months of demand on 105,000 electronics items. The performance criteria examined were the percent error and mean squared error (MSE) over the items' leadtimes. The results showed an overall average forecast error of 144%. The best alpha for quarterly forecasted items was shown to be .2, while for monthly forecasted items (VIP items) the best alpha was .01 (although as the author points out, this would result in virtually constant forecasts from month to month). The study also suggested that some improvement in forecast accuracy can be obtained using longer forecast intervals: that is, forecasting quarterly items semi-annually, and forecasting monthly items quarterly. The study of methods for improved forecasting is continuing at the DESC operations research (OR) office, and similar work is underway at the Defense Construction Supply Center (DCSC) OR office as well.

Another interim report described the preliminary results of a study of subsistence items (7). The study examined 77 months of data for 3,940 items. Various seasonal and nonseasonal autoregressive (AR) models were examined, along with the current weighted average method, and 6- and 12-month moving averages. The results showed that five models proved to be good performers for about 75% of the items: AR1, AR1 seasonal, AR1 trig seasonal, and the 6- and 12-month moving averages.

A recently completed study examined the Program Oriented Item (POI) system, used by the Defense Personnel Support Center (DPSC) to forecast some of their clothing items (8). The study did not compare forecasting methods, but did identify POI items which had seasonal and trend components. The study showed that moving averages outperformed the POI system for trend items, and that Winters' triple exponential smoothing performed better for seasonal items.

To summarize, most of the studies of DLA forecasting seem to confirm that exponential smoothing with alpha values of around .2 is a superior method to other simple approaches. Some of the studies' findings suggest that longer moving averages, such as 4-quarter, might be effective for some items. Results for subsistence and POI items suggest that alternate models, such as Winters method or autoregressive models, might be preferable.

Both of these latter studies, however, examined commodities which might be expected to have a higher proportion of items with trend or seasonality. Finally, DESC's results suggest that monthly forecasting, while perhaps useful for management purposes, may not improve forecast accuracy. All of these findings will be taken into consideration in the evaluation of methods for inclusion in the present study.

C. Forecasting Efforts in the Services

The Army, Air Force and Navy have all produced many studies related to their forecasting systems. Each of these Services was contacted by DLA-LO, and many of their forecasting studies were reviewed for this effort. There is, however, at least one major difference between the Services and DLA in terms of forecasting. The Services manage both reparable and consumable items, while DLA manages only the latter. This factor has led the Services to the use of program factors, such as flying hours for aircraft, to forecast demand for some (reparable) items. With the exception of the POI system, DLA generally does not use program factors. Therefore, at least some of the work done by the Services is not directly applicable to DLA. There are, however, two efforts which are particularly relevant and will be discussed here.

The Operations Research Analysis Department of the Navy Fleet Material Support Office (FMSO) is currently involved in a forecasting study quite similar in nature to DLA-LO's study. A meeting was held with the analysts at FMSO to discuss the two studies and share information regarding forecasting.

The other effort of particular relevance to the current research is the forecasting study completed for the Services and DLA by Boeing Computer Services (9). The study was originally intended to use data from all services, but ended up examining data from the Army and Navy. A total of 60 quarters of data for 23,911 Army items, including program data (flying hours), was included in the study. The Navy data consisted of 36 quarters of demand for a sample of 900 items. A series of forecasting methods was examined for each service, including naive methods, exponential smoothing, linear regression (using flying hours), 8-quarter moving average, two Autoregressive Integrated Moving Average (ARIMA) models, and a method developed by Steece and Wood which combines items into groups in order to generate forecasts.

The results of the study showed that the 8-quarter moving average was the best of the simpler methods. Exponential smoothing did not perform well in this study; the regression using program factors was also a poor performer. The AR(1) model, however, was judged to be a good performer. The results also suggested that the Steece-Wood method may be a useful one, provided meaningful item groupings can be determined. These results will also be considered in evaluation of methods for the current study.

D. Academic Literature

Voluminous academic literature exists concerning forecasting, and a review of this literature would be prohibitive. However, several studies are worthy of discussion here, either because of their scope in comparing forecasting methods, or because they represent summaries of specific areas of the literature.

The most comprehensive comparative study of forecasting methods developed as a result of a forecasting "competition" conducted by Makridakis (10). Experts in the field applied approximately 20 different forecasting methods to 1,001 different time series. The data series were monthly, quarterly, and yearly, and consisted of micro-level data (for an individual company, for example) or macro-level data (GNP, for example). Various forecasting horizons were examined and several different error statistics were calculated. Although the results varied depending on the nature of the series examined, some general conclusions are offered by Makridakis et al. First, it is not necessarily the case that complex methods produce more accurate forecasts than simple methods. According to the authors, the more noise or randomness in the data, the less important it is to use sophisticated methods (10, p. 127). In addition, the study showed that deseasonalizing the data (that is, removing seasonality) using simple decomposition techniques is adequate, and produces similar performance among most of the forecasting methods. Finally, the results showed that single and double exponential smoothing, applied to deseasonalized data, do well for short forecast horizons (1-2 periods ahead), while the Holt, Brown and Holt-Winters double exponential smoothing methods do well for forecasts 3-6 periods ahead.

One other result from the Makridakis study is notable. The methods which combined forecasts performed very well in the study. The combined forecast always outperformed its individual components. This finding, as well as the others, is discussed further in the next section of the report.

In addition to this important study by Makridakis, several very useful survey articles have appeared recently. One of these is by Armstrong (11), who seeks to summarize the results of previous research on forecasting methods such as those discussed up to this point. The first conclusion offered by Armstrong echoes Makridakis': sophisticated methods seem to perform no better than simpler methods. In fact, Armstrong suggests that when limited historical data are avialable, highly complex models may actually serve to reduce forecast accuracy (11, p.55). Another conclusion was that combining forecasts seems to be a promising approach to improving forecast accuracy. Armstrong does point out that little evidence is currently available regarding the best way to weight the components of the combined forecasts.

Mahmoud (12) reaches much the same conclusions in his survey of the forecasting literature. Reviewing some 100 forecasting studies, he too concludes that simple forecasting methods perform as well or better than more complex methods (12, p. 153). He also notes that several studies show that exponential smoothing performs better over a relatively short-term forecasting horizon (less than one year) than over a longer period. Mahmoud also concludes, along with Makridakis and Armstrong, that combining forecasting results produces better forecasts (p. 154).

Finally, a recent article by Gardner (13) reviews and summarizes the literature on exponential smoothing. In addition to providing exponential smoothing models for seasonal and trend series, Gardner also offers some conclusions about the specifics of exponential smoothing. He suggests that parameters for the models should be estimated from the data, and not pre-selected. He does note, however, that moderate parameters, say .2 or .3, are appropriate in inventory applications where forecasts are generated automatically (p. 11). Gardner goes on to point out that although linear trends are usually used in exponential smoothing models, there is evidence that the trend should be "damped" (i.e., slowed) as the forecast horizon increases. Finally, the article notes that there is no strong evidence suggesting the superiority of adaptive smoothing methods, which allow the alpha values to change from one period to the next, over standard exponential smoothing methods which do not allow the alpha value to change.

To summarize, these survey studies of forecasting methods seem to agree that simple forecasting techniques, such as exponential smoothing, perform as well as or better than complex techniques, such as ARIMA, for many applications. They also seem to agree that the method of combining forecasts holds much promise for improving forecast accuracy. Both of these conclusions will be considered in the next section of this report, which addresses the selection of models for inclusion in the present study.

III. REVIEW AND EVALUATION OF FORECASTING METHODS

A. Introduction

One of the purposes of the literature review was to identify forecasting techniques for possible inclusion in the present study. This process consisted of two phases. Initially, any method identified in the literature was considered. Descriptions of each technique were developed and these were then reviewed by all project staff. The relative merits of each method were considered, and a judgment was made regarding the inclusion of each method in the study. The methods were judged based on their applicability to DLA's forecasting needs, their anticipated accuracy based on previous studies, and the cost associated with their implementation and maintenance. This last consideration is obviously an important one, since DLA must forecast a large number of items each quarter.

This section presents a brief discussion of each forecasting method, along with the reasons for including or excluding each from the study.

B. Forecasting Methods Examined

1. Moving Average

The moving average (MA) technique uses the arithmetic mean of the last "n" periods as the forecast for the next period. The advantage of this method is its simplicity. The disadvantages are (a) it will not successfully forecast seasonal data, and (b) the forecasts will lag behind any trends in the data. Deseasonalizing the data can circumvent the first problem. Due to its ease of use, the MA method was included in the study. Both 4-period and 8-period moving averages were examined.

2. Single Exponential Smoothing

Single exponential smoothing (SES) uses a constant value (alpha) to "smooth" the current observation; the larger the value of alpha, the greater the weight given to the current observation. The forecast consists of the weighted current observation plus the previous smoothed value of the series.

The advantage of SES is its ease of implementation; it requires fewer data points to store than the moving average. The major disadvantages of SES are the same as those of the MA method. Due to its ease of use, SES was also included in the study.

3. Brown's Double Exponential Smoothing

Brown's double exponential smoothing (DES) uses two smoothing equations; one to smooth the current observation, and a second to smooth the smoothed value of the first equation. The method also uses the difference between the two smoothed values as a measure of trend in the data.

The advantages of DES are the same as those for SES, with the additional fact that DES can forecast trends in the data. The disadvantage is that Brown's DES does not allow for seasonality in the data. Again, deseasonalizing the data can correct for this shortcoming.

Since a version of Brown's DES is currently used in DLA, this method was included in the study. The version of Brown's method used in SAMMS (without the trend term) was included in the study as well.

4. Holt's Double Exponential Smoothing

Holt's version of DES represents an alternative to Brown's formulation. The Holt method differs from Brown's DES in that it uses two smoothing parameters rather than one. The level of the series is obtained by using alpha to smooth the current observation into the previous level plus trend terms. The trend term is obtained by using gamma to smooth the difference between the current and previous levels into the previous trend term.

The Holt method has the potential advantage of increased accuracy associated with the use of multiple smoothing factors. This can also be a disadvantage, however, since values must be chosen and maintained for two constants, rather than one. The Holt method was included in the study in order to compare it with Brown's DES.

5. Gardner's Double Exponential Smoothing

This method is a variant of the Holt procedure developed in recent work by Gardner. (14) Gardner proposes applying a third smoothing term, phi, to Holt's equations. The phi parameter is applied to the trend term, and would usually range from 0 to 1. If phi is 0, the model is equivalent to simple smoothing. If phi is 1, the model is the same as Holt's model. If phi is between 0 and 1, however, then Gardner's method "damps" the trend; that is, the trend is assumed to change at a slower rate than is implied by the Holt model (14, p. 5). This technique was included in order to compare it to the standard Holt procedure.

6. Adaptive Exponential Smoothing

Adaptive exponential smoothing (AES) methods allow the value of alpha to change as patterns in the data change. Four adaptive smoothing techniques were considered for inclusion in the study. These techniques use different methods to adjust the smoothing constant, depending on the error being produced by the current constant. The ideal AES method should be responsive to changes in the data, and yet should not be overly sensitive to large, one-time fluctuations.

Two of the AES methods, one developed by Eilon and Elmaleh (15) and one by Roberts and Reed (16), use a periodic review technique; that is, the smoothing constant is reviewed for change only after several periods have passed. The periodic review technique is considered to be relatively unresponsive to changes in demand. Therefore, neither of the AES methods were included in the study.

The first of the continuous review methods examined was the Whybark (17) method. The Whybark method allows for specification of three values of the smoothing constant which allows the forecast to be adjusted more quickly when the forecasts move away from the observations. While this method would work quite well with relatively clean data, the noise anticipated with the data used in this project would require some sort of filter to prescreen the data. This would increase the computation involved and dilute the responsiveness of the method. Due to these factors, this AES method was excluded from further consideration.

The method found to be most promising is the one proposed by Trigg and Leach (18). This method adjusts the smoothing constant each period based on the ratio of the smoothed error to the smoothed absolute error. The use of smoothed error terms in the tracking signal allows the forecaster to have some control over the sensitivity of the signal to the last error term in the series. The method should be quite responsive, since the signal is adjusted after each observation. Given these factors, Trigg-Leach was the AES method included in the study.

7. Decomposition

Decomposition methods are used primarily to identify seasonal factors which can be used to remove seasonal variations from the data. Forecasting methods which cannot handle seasonal data can then be applied to the deseasonalized data. The most basic approach is known as the ratio-to-moving averages classical decomposition method, which uses a moving average to deseasonalize the data.

Classical decomposition seems better suited to the purposes of this project than the main alternative, known as the Census II method. This latter method is much too complex and involved to be implemented in the present context. Therefore, the classical method was used in the present study, with a 4-quarter moving average as the basis for deseasonalization.

8. Autoregressive Integrated Moving Average

The term "autoregressive integrated moving average", or ARIMA models, was popularized by Box and Jenkins (19). Basically, autoregressive models base their forecasts on equations which differentially weight each of the previous observations. Moving average models use previous error terms associated with past observations to derive a forecast. ARIMA models combine autoregressive (AR) and moving average (MA) models.

ARIMA models are usually associated with the time series analysis process described by Box and Jenkins (19). This is a three-step iterative process which involves model identification, parameter estimation, and forecasting. The first two of these steps are rather involved, and would require automatic methods to handle the number of items involved in the current application. It is possible, though not necessarily desirable, to skip the model identification step and simply apply one or more ARIMA models to the data. This still involves a rather lengthy coding process required in order to develop parameter estimates. Given all this, it was decided to perform a test on a limited number of items, using the SPSSx statistical package's Box Jenkins procedure, to determine whether the benefits outweighed the disadvantages described above.

A total of 100 items were selected at random from the larger sample. Only those items which had the maximum amount of data (eight years, or 32 quarters) were included in the 100 examined.

The first step in identifying the models was to analyze the plots of the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) of each of the time series. The ACF is a measure of the relationship (correlation) of the time series with itself, lagged by some number of time periods. For example, the autocorrelation for the first lag measures the relationship between the demand at each time period and the demand at the time period preceeding it. The autocorrelation at the second lag measures the relationship between demand values two periods apart. In general, the ACF at the k^{th} lag measures the relationship between observations k periods apart. The PACF is a measure of the degree of association between the series and the k^{th} lag of the series when the effects of the other time lags are partialled out. The PACF is used in conjunction with the ACF to identify an appropriate ARIMA model for forecasting.

The results of this analysis showed that 80 of the 100 series had no useable pattern to the autocorrelations; that is, the demand data was essentially random. There are two likely reasons for this finding. First, 32 quarters is a small number of data points on which to base such an analysis. In addition, many of the series contained a large proportion of zeros, making the identification of a reasonable model more difficult.

Of the remaining 20 time series, 18 were fit to nine different ARMA models (two series could not be satisfactorily fitted to any model). AR(1) models were selected for five items, and ARMA(1,1) models were fitted to four items. The remaining items were distributed over the other seven models.

The results of the test analysis presented above were not very encouraging. Under the best of circumstances (a full 32 quarters of data), models could be identified for only 18% of the items. Although a single model (AR1, for example), could be used to forecast all items, the decision of the project staff was that the time required to develop the code for the method would not be offset by the potential benefits. Therefore, ARIMA models were not included as a part of the study.

9. <u>Steece-Wood</u>

The Steece-Wood method involves using a complex model, such as an ARIMA, to forecast an aggregate series, and a simpler model, such as exponential smoothing, for the series that comprise the aggregate (20). The success of the method appears to depend on the ability to develop meaningful aggregate series. Although there would be various ways to divide DLA's items into groups, none of these would be likely to produce aggregates which could be forecasted more successfully than the individual items themselves. Therefore, due to the inability to form meaningful aggregates, the Steece-Wood method was not included in the study.

10. Transfer Function Models

Transfer functions are used in Box-Jenkins ARIMA models in order to analyze multiple time series; that is, to forecast several related series simultaneously. One way in which this method could be applied to DLA is in the use of program factor data, in addition to the historic demand, to forecast future demand. Program factors are not readily available for DLA data, and are of questionable utility in forecasting DLA's items in any case. This method was therefore not considered further in the study.

11. Econometric Models

These models were not considered to be relevant to DLA's data. The models are difficult and costly to develop and maintain, and usually involve some underlying theory regarding system functioning. These models were not included in the study.

12. Regression Models

The main utility of regression models is to relate independent variables, such as program factors, to the series to be forecasted. Since program data is believed to be of limited use here, regression models could not be used for this purpose. In addition, examination of the relationship between the available item characteristics and demand quantity failed to reveal any convincing evidence for the use of these variables in forecasting demand. Moreover, any of these external variables would themselves need to be forecasted prior to developing the demand forecast, and the error associated with the former forecasts would increase the error in the latter forecasts. Given these considerations, regression models were excluded from further consideration.

13. Kalman Filters

The Kalman filter is a dynamic linear system where the system of equations specifies: (a) how observations of a process are stochastically dependent on the current process parameters, and (b) how the process parameters evolve in time, both as a result of the inherent process dynamics and from random disturbances. The use of time-varying coefficients allows the forecasting model to adapt over time. This flexibility increases forecast accuracy through continuous reestimation of parameters as new observations become available. In addition, Kalman filters can be used to detect significant changes in the time series and to adapt to these changes.

There are several variations of the Kalman filter currently in use. For example, if the criterion is to minimize the future MSE of a model fitted to historical data, Kalman filters do as well, if not better, than classical estimation procedures.

Since the Kalman filter is an extremely complex model to implement, it was decided to seek the advice of forecasters experienced with this particular method. Comments from forecasting experts in both the government and private sectors were obtained. The general concensus was that the Kalman filter, given its use of time-varying coefficients, is inefficient for forecasting demand for a large inventory of items. This is due to the large amount of noise inherent in the demand streams for many inventory items, noise which is more severe than is found in the types of engineering applications for which the filter was designed. Due to these considerations, the Kalman filter technique was not examined in the study.

14. Forecasting of Leadtime Demand

This area was selected for review as a specific example of the more general area of distribution fitting for forecasting. The method of forecasting leadtime demand was the only one which actually detailed the process of identifying and fitting a distribution to the demand series, and then generating a forecast based on this distribution. This specific method, however, emphasized setting a reorder level, as opposed to a demand forecast. That is, the method is concerned with minimizing the risk of stockout conditions, which is only one specific criterion by which to judge a forecasting method. Therefore, this method was excluded from further consideration.

15. Focus Forecasting

Introduced by Smith (21) focus forecasting begins with a number of simple or "common sense" models. Each period, all models are tested to determine which would have best forecasted last period's demand. The model selected is then used to forecast next period's demand. The main problems with the method include its overly simplistic approach to forecasting, and the total lack of empirical evidence in support of the approach. It was therefore rejected from further consideration in the study.

16. Combining/Averaging Methods

There appears to be good support in the forecasting literature for the usefulness of this forecasting approach (see, for example, ref. 22). Several different methods for combining forecasts were examined for inclusion in the study. The most basic is to take a simple average of the forecasts produced by each method being combined, and use this as the forecast for the next period. Makridakis et al. (10) found this method to be a very successful one.

An alternative to using a simple average is to weight each forecast according to the past error associated with each. Brandon and Lackman (23) present evidence for the usefulness of this type of method. Their procedure takes into account both the mean squared error (MSE) and the standard deviation of the errors (SDE) for each forecasting method. Each method's weight is represented by $1-p_a$, where p_a is the ratio of the error produced by method 'a' to the overall error produced by all the methods combined. The calculation of the MSE and the SDE can include as many historic time periods as is deemed appropriate by the forecaster.

In the weighting scheme proposed by Brandon and Lackman, the weights for the different forecasting methods are forced to sum to one. A recent article by Granger and Ramanathan (24) suggests that this is not necessarily the best combination procedure. These authors begin with a discussion of linear combinations of forecasts where the weights, obtained using least squares, are constrained to sum to one. They go on to demonstrate the superiority of a method which does not restrict the weights, but which does add a constant term to the least squares formulations.

While Granger and Ramanathan may make a convincing argument, their method of combining forecasts appears much too involved and complex to be of practical use in meeting DLA's forecasting needs. Brandon and Lackman's method, however, appears relatively easy to implement, and has intuitive appeal as well. This method was therefore the combination method used in the present study.

In addition, since there is very little additional cost or effort associated with the simple averaging method, this technique was also included in the study.

17. _"Naive" Methods_

In addition to the methods identified above, there are a number of relatively simple or "naive" methods which were considered. Several of these methods were shown in the literature review to work well in various applications. The main advantage of such methods, however, is that they are quite easy to develop and implement, and are relatively cost-free to maintain. Due to this consideration, the following "naive" models were included in the study:

- naive forecast (last period's demand)
- simple mean of past data

Both of the above methods were applied to both the original data and the deseasonalized data.

C. Summary

This section reported the results of the screening process designed to determine those forecasting techniques to be included in the current study. Based on the applicability of each method to DLA and the cost associated with implementing and maintaining the method, each technique was included or excluded from the study. The results of this process resulted in the following techniques being included in the study:

- naive forecast (last period's observation)
- simple mean of past observations
- 4-period moving average
- 8-period moving average
- single exponential smoothing
- current DLA version of exponential smoothing
- Brown's double exponential smoothing
- Holt's double exponential smoothing
- Gardner's double exponential smoothing
- Trigg-Leach adaptive exponential smoothing
- a combination/average of two or more of the above

The formulae for these methods are shown in Appendix A. In addition, eight of these methods (the first seven listed plus Trigg-Leach) were examined using both the raw data and the data after it had been deseasonalized using the ratio-to-moving averages classical decomposition method.

IV. METHODOLOGY

This section provides a description of the development of the data and procedures used in the study. It is divided into three subsections, which describe (a) the development of the data base for the study, (b) the selection and validation of the item samples used in the study, and (c) the procedures used to implement the forecasting methods identified in Section III of the report.

A. Development of Study Data Base

1. Demand Data

There are several factors related to the study which guided the initial search for data sources. The study required as much historic data as were available for all DLA commodities. In addition, it was desirable to be able to segregate types of demand: that is, recurring versus non-recurring, Foreign Military Sales (FMS), and Government Furnished Material (GFM).

After examining available data sources, it was determined that historic Supply Control Files were the best source. These files contain, for items which are family heads, demand by quarter, by type of demand (recurring/non-recurring), and by source of demand (GFM, FMS). They also contain information describing the item; that is, the various item characteristics (e.g., supply status code, weapon system code) required for the later phases of the study.

The study team was able to assemble a collection of historic Supply Control Files. In general, the Supply Control Files were available for the time period beginning with fiscal year 1977 and ending with fiscal year 1984. Several files were missing, however, for individual commodities. Extensive efforts to locate these missing files met with no success. A summary of the data available for the study is presented in Table 1.

Table 1

SUPPLY CONTROL FILE DATA AVAILABLE

Commodity	Fiscal Years	Maximum Quarters of Continuous Data
Construction	1977-1984	32
Electrical	1977-1984	32
General	1981-1984	16*
Industrial	1977-1984	32
Medical	1977-1984	32
Clothing & Textiles	1982-1984	12**

*Fiscal years 1979 and 1980 were missing for General. Since continuous data is required, General begins with FY 1981.

**C&T was not on SAMMS prior to FY 1982.

The next step in the construction of the data base was to determine which items should be included. The decision was made to identify those items which were actually forecasted by SAMMS at the point in time where the data available for the study ended (i.e., 30 September 1984). It was felt that this approach would most closely simulate the actual current forecasting situation. For example, DLA must currently forecast items which have varying amounts of historic data. It would therefore be unrealistic to include in the present study only those items with a full 32 quarters of demand data. Instead, the study will seek to identify forecasting techniques which will be successful for all of the items DLA must forecast.

The criteria used to select the items to be included in the data base were those currently used by SAMMS to determine which items receive the exponential smoothing forecast. These criteria are:

1. Demand supported replenishment items (Item Category Code '1')

2. Established items--usually those over two years in DLA (Age of Item Code 'E')

3. Stocked items (Supply Status Codes other than '2', '3', '9')

An item which passed these three criteria would be forecasted by

SAMMS and should be included in this study.

A total of 677,705 items met these three criteria. These 677,705 items became the total population of items to be included in the study. At this point, each of these items was matched back to past Supply Control Files, and all of the historic data available for that item were obtained. In order to allow the greatest amount of flexibility in the data base, the frequency and quantity of demand were obtained by quarter separately for each of the following types of demand: recurring, non-recurring, GFM, and FMS. The results of this process showed, as expected, varying amounts of past data available for the population of items. Table 2 shows the number of quarters of historical data available for the items in each commodity.

Table 2

NUMBER OF QUARTERS OF DATA AVAILABLE

	Commodity					
Qtrs	C	E	G	I	M	T
32	60,160	68,666	0	215,506	9,469	0
28	2,720	67,964	0	8,970	344	0
24	3,883	7.027	0	11,215	344	0
20	4,134	6,030	0	7,977	1,181	0
16	4,767	7,315	77,780	7,453	176	0
12	17,497	12,503	8,199	30,571	304	14,001
8	1,871	3,429	1,493	7,137	576	527
4	857	1,588	582	3,171	136	182
Total	95,889	174,522	88,054	292,000	12,530	14,710

Note. Commodity abbreviations are as follows: C = Construction, E = Electronics, G = General, I = Industrial, M = Medical, T = Clothing and Textiles.

Two final points regarding the data base should be noted here. First, the supply control files contain data only for items which are family heads. If an item switched from family head to family member during the time period examined, the demand for the item as a family member would not be included in the study. Demand for family members is "rolled up" to the family head. It should also be noted that the Supply Control Files used to build the data base are collected after fiscal year-end processing (i.e., as of 1 October).

The other point relates to the item characteristic data mentioned previously. All data were obtained from the 30 September 1984 Supply Control File. No effort was made to track changes in the item's characteristics over the years under consideration. This approach is consistent with the idea of simulating the current information available to the system for forecasting.

To summarize, Supply Control Files from FY 1977 thru FY 1984 were used to obtain historic demand data for the study. A total of 677,705 items which were forecasted by SAMMS on September 30, 1984 will serve as the population of items to be used in the study. Quarterly data was collected for various categories of demand: recurring vs nonrecurring, FMS, and GFM.

2. Item Characteristics

As noted previously, one of the goals of the present study is to match item characteristics with forecast accuracy to attempt to determine which forecasting methods work best for which kinds of items. This section will discuss the item characteristics which were available on the FY 1984 Supply Control File, and will evaluate these in terms of their usefulness in accomplishing this goal.

A total of 35 variables or item characteristics was obtained from the Supply Control File. Each of these variables was examined by the study team to determine whether it would be a useful one to attempt to relate to forecast accuracy.

A variable was not included in the study for one of three reasons. First, any variable which was directly related to the current forecasting method was not considered appropriate for inclusion in the study. If alternate forecasting methods were recommended, variables which relate to the exponential smoothing method performed by SAMMS would not be available for these items. This criterion eliminated the following item variables from further consideration: QFD, new QFD, demand value code (based on QFD), single smoothing constant, double smoothing constant, procurement cycle, safety level quantity, sum of forecast errors, mean absolute deviation of forecast errors, alpha factor, out-of-track signal, and forecast basis code.

A second reason for eliminating a variable from the study relates to the distribution of items over the categories of the variable. A variable for which a very large percentage of items have the same value is not very useful in the current context. As an example, age of item code and item category code would not be useful variables in this study since all items have the same values for these variables ('E' and '1', respectively). This second criterion resulted in the elimination of the following variables: method of computation code (98% blank), future supply status code (97% 'N' for 'No Change'), VIP code (96% 'N' for 'Non-VIP', indicating quarterly forecast), and shelf life code (98% '0', indicating no shelf life restrictions).

The third reason for excluding variables from consideration is a "logical" or "common-sense" one. There are some variables which would simply not be expected to be related to the ability to forecast demand for an item. Such variables should not be included in the study, since they may lead to spurious findings. Based on the best judgment of the study team, the following variables were eliminated for this third reason: months since management assumed (based on last buy date), administrative lead time, production lead time, fixed safety level, operating level, annual non-recurring demand percentage, and storage mission code. One final variable, essentiality item code, is assigned by individual Supply Center and has no common meaning from center to center; it was, therefore, excluded from further consideration.

This screening process left a relatively small number of variables to work with. The first of these was supply status code. Although only three categories ('1', 'A', and '6') accounted for 99% of the items in the population, this variable was felt to have enough potential usefulness to be included in the study.

The second variable included in the study was months since system entry, which is defined as the number of months between the date of system entry and September 1984. This variable is an indicator of the level of activity of the item, which is expected to be related to the ability of different forecasting methods to accurately predict demand.

One final variable included in the study was the weapon system indicator code. About half of the items in the population were non-weapon system items. Due to the recent increased emphasis on weapon systems support within DLA, this variable was included in the study.

To summarize, 35 item characteristics were initially examined for inclusion in the study. Most of these variables were rejected due to (1) their relationship to the current forecasting system, (2) their inability to differentiate between items, or (3) the lack of a logical basis for their being related to demand forecasting success. The variables which will be included in the study are: supply status code, months since system entry, months since last demand, commodity code, and weapon system indicator code.

In addition to these item characteristics, several variables were created based on historical demand. These variables are:

- number of quarters of data used in forecast
- recurring demand quantity for last year

- recurring demand frequency for last year
- nonrecurring demand quantity for last year
- nonrecurring demand frequency for last year
- percentage of quarters with zero demand
- mean demand quantity for all available quarters
- standard deviation (SD) of demand
- mean of first differences of demand for all quarters divided by mean demand quantity
- standard deviation of first differences of demand divided by mean demand quantity
- percentage of demand three SDs above or below the mean

The first eight variables listed above require no additional information. The next two variables are based on the first differences of the demand series. The first difference is obtained by subtracting the demand for each quarter from the demand for the subsequent quarter (e.g., demand for Quarter 2 - demand for Quarter 1). The size of the mean of the first differences of the demand is an indication of the amount of fluctuation in the demand data (the larger the mean, the larger the variability in the data). In addition, a positive mean indicates that demands are increasing in size over the time period under consideration (upward trend), while a negative mean indicates the opposite (downward trend). The SD of the first differences is an indication of the regularity of these trends in the data. Both variables are divided by the average demand size, providing a relative measure of the change from one quarter to the next quarter.

B. Sample Selection

The population to be included in the present study consists of 677,705 items which were forecasted by DLA as of 1 October 1984. It is obviously not desirable for the present study to compare all forecasting techniques for such a large number of items. It was therefore necessary that a sample of the total number of items be obtained for use in the study. Additional samples are also required to verify the results obtained using the first sample.

The basic concept behind sampling is quite simple. If the goal is to draw conclusions about some population (like the 677,705 items in this study), it is not necessary to examine the entire population. Rather, a smaller group of representative items can be selected for study. So long as the sample items are representative of the population as a whole, there is reasonable confidence that any findings which hold for the sample will apply to the entire population as well. The most effective method for ensuring the representativeness of the sample is to draw members at random from the population. A completely random sample should be representative of the population from which it was drawn.

It was determined that 1% samples of the population would be selected for study. The resulting number of items would be large enough to adequately represent the entire population, yet is a manageable size for comparing the various forecasting methods.

Three 1% samples were selected randomly from the population of items. The three samples were selected in turn, and any item included in one sample was excluded from subsequent samples. The sample sizes were 6,829 items, 6,815 items, and 6,499 items, respectively, for the three samples.

Appendix B presents a comparison of the population and three samples on three variables: commodity (Table B-1), supply status code (Table B-2), and number of quarters of demand data available (Table B-3).

Table B-1 presents the distributions of the sample and population items by commodity. In general, all samples appear to be representative of the population on this variable. Sample 1 has a slightly greater proportion of items in Construction and a slightly lower proportion in General, when compared with the entire population. Sample 2 has a slightly higher percentage of items in Industrial, and slightly lower percentages in Construction and Electronics. Finally, Sample 3 has a larger proportion of Industrial items, and a lower percentage of Electronics items, than the population. None of the above differences are considered significant.

Table B-2 shows the supply status codes (SCC) for the population and three samples. Sample 1 has a slightly larger proportion of SSC "1" items than the population. Aside from this, there are no apparent differences among the three samples and the population on this variable. Table B-3 shows the amount of data available for the items in the population and samples. Sample 3 has a slightly larger proportion of items with all 32 quarters of demand data. Otherwise, there were no significant differences between the population and samples.

In summary, three random 1% samples were drawn from the population. A comparison of selected characteristics showed that all samples appeared to be representative of the population as a whole. Sample 1 was used to conduct the preliminary analyses involved in the study. Samples 2 and 3 were used to validate the findings obtained using Sample 1.

C. Forecasting Procedures

The basic procedure followed in the study was to forecast each item in Sample 1 with each of the 18 forecasting methods. The items had from 4 to 32 quarters of data available. The procedure followed in all cases was to withhold the four most recent quarters of data to assess the accuracy of the model, and use the remaining data to fit the model. This meant that those items with only four quarters of data were eliminated from the study (59 items, or 0.8% of the sample, were eliminated for this reason). Also, any item which had zero demands in all but the last four periods was excluded from further study (358 items, or 5.3% of the sample, were excluded for this reason). These exclusions left a total of 6,412 items remaining in the sample.

All of the exponential smoothing methods require the use of at least one smoothing parameter. For each of these methods, individual parameters were found for each item. This was accomplished by testing 11 different values (0 to 1 by increments of .1) for each parameter. The value which produced the smallest root mean square error (RMSE) for the one-period ahead forecasts for all periods was the one used to forecast that particular item. In the cases of the Holt and Gardner methods, 121 and 1,331 parameter combinations (respectively) were tested for each item. A listing of the various parameters' frequency of occurence is shown in Appendix C.

The SAMMS version of double exponential smoothing was included in the study as a baseline against which other methods could be compared. For purposes of comparison with other methods, individual parameters were calculated for each item. In addition to the basic formulas, SAMMS takes two additional actions in its computation of the forecasts which were included in this method's calculations. First, any forecast which was less than 1 was set equal to 1. Second, if a forecast was negative, it was replaced by the average of the two most recent quarters of demand, as were the single and double smoothed averages.

The decomposition of the data was accomplished using the ratio-to-moving averages classical decomposition method as described in Makridakis (25). This method involves replacing each raw data point with a centered 4-quarter moving average. The resulting values are free of annual seasonal influences. The method goes on to derive seasonal factors for each quarter, which are based on the proportion of each quarter's demand to the overall demand for each year. The deseasonalized data stream is then forecasted using one of the eight methods described earlier. The resulting forecasts are multiplied by the corresponding seasonal factor in order to arrive at the final forecast.

It should be noted that the decomposition procedure results in the loss of three data points: two at the beginning of the series, and the final point in the series (this is due to the fact that the moving average is centered).

The two moving average methods are the only ones which are affected by the loss of additional data points when the data are deseasonalized. For the 4-quarter moving average, items with only eight quarters of data could not be forecasted, while for the 8-quarter moving average, items with 12 or fewer data points could not be forecasted. This meant that these methods were applied to fewer items than the other methods. Specifically, 86 items could not be forecasted using the 4-quarter moving average with the decomposed data, and 797 items could not be forecasted using the 8-quarter moving average.

All of the exponential smoothing methods employed backcasting in order to determine initial values for the key terms in the equations. Backcasting, a technique introduced by Box and Jenkins (19), involves reversing the order of the data in the series, and applying the forecasting method to the reversed data. The values for terms at time zero are then used as initial values in the actual forecasting procedure.

As noted previously, forecast accuracy was assessed over the last four periods only. For each method, two different forecasts were generated. The first was a short-term, or one-step ahead forecast. This forecast uses the actual data from the most recent past period to compute the forecast for the next period. For example, in single exponential smoothing, the one-step ahead forecast for the 32nd quarter (assuming 32 quarters of data) would include the actual demand up to and including the 31st quarter.

The other forecast generated will be referred to as the long-term forecast. In this forecast, it is assumed that we are currently at period 28 (again assuming 32 quarters of data) and must forecast the demand for periods 29-32. Since the demands for these latter periods are unknown, they cannot be included in the forecast. It should be noted that for the methods which fail to take trend or seasonality into account (such as simple exponential smoothing), the four long-term forecasts will all be equal to the first one-step ahead forecast. This is not the case for methods which do take trend into account. Double exponential smoothing, for example, multiplies its trend term by the number of periods ahead to be forecasted, resulting in a different long-term forecast for each of the four withheld periods.

Appendix D provides an example of the two forecasts generated. The appendix illustrates the two approaches for both single exponential smoothing and double exponential smoothing.

In the application of forecasting to the inventory environment, the forecasts must be made over a long horizon (for DLA, the length of the leadtime plus the procurement cycle). Obviously, there is no information available concerning the demand for subsequent time periods. Therefore, it would appear that the long-term forecasts, and the error associated with these forecasts, are a more appropriate measure for use in the present study.

The measure of forecast accuracy used in the study is the root mean square error, (RMSE) as described by Armstrong (26). The formula for the RMSE is:

$$RMSE = \frac{\left[\sum_{t=1}^{n} (X_t - F_t)^2\right]^{\frac{1}{2}}}{n}$$

where X_t is the actual demand for time period t
F_t is the forecast for time period t
n is the number of time periods over which the error is calculated (n=4 in the analyses presented here).

The RMSE produces large penalties for large forecast errors by squaring the error term in the numerator; otherwise, it is similar to the mean absolute deviation (MAD; see reference 26, p. 321).

V. ANALYSIS

A. Results of Preliminary Analyses

1. Introduction

This section describes the results of three preliminary analyses conducted on the first sample. The goal of these analyses was to compare all of the forecasting methods identified previously in order to eliminate from further consideration those methods which were poor performers.

Prior to the comparative analyses, a first step in the analysis was to attempt to get some indication of the existence of trend, seasonality, and randomness in the demand data streams for the items in the sample. This was done by computing autocorrelations for the first four lags. These were then compared to a 95% confidence interval obtained by multiplying plus or minus 1.96 by the estimate of the standard error (the standard error estimate is $1/\sqrt{N}$, where N is the total number of data points in the series). If none of the first four autocorrelations was outside this interval, then the item's demand series was considered to be random (that is, having no identifiable pattern). If the fourth autocorrelation only was outside the interval, the data stream was considered to be seasonal. If the first three or all four autocorrelations were outside the interval, the demand contained trend. Finally, if one or more of the four autocorrelations were outside the limits (other than the fourth), the series was assumed to contain some pattern other than trend or seasonal.

Based on the procedure described above, 79% of the items in the sample were judged to be random (no identifiable pattern). An additional 1.9% had trend in their demand streams, and 1.5% had seasonal demands. The remaining 17.5% of the items had some pattern to the demand other than trend or seasonal.

These results will be helpful in the comparison of the forecasting methods to be presented in this section. It should be noted, however, that the use of the autocorrelations as described above is a convenient but weak indicator of the existence of trend and seasonal patterns.

2. First Preliminary Analysis

Table 3 presents the long-term RMSEs for all 18 forecasting methods. The table shows the average and median RMSEs, and the sample standard deviation of the errors, for the 6,412 items sampled (it should be recalled that the two moving average methods which used the decomposed data were not applied to all items in the sample). The table also shows the ranks of each method based on the median and the standard deviation of that method. The methods in the table are listed in order of increasing average RMSEs.

As the table shows, the 8-quarter moving average (MA) applied to the deseasonalized data (Dec MA8) had the lowest average RMSE score. As noted previously, however, this method could not be used for 797 (12.4%) of the items in the sample. The number one ranking of this method fails to take this into account, and must, therefore, be viewed with some caution.

The study's approximation of the current SAMMS forecasting method ranked fifth overall. Single exponential smoothing (SES) and the two moving average methods, in addition to the decomposed moving average, performed better than the SAMMS version of double exponential smoothing.

The sample standard deviation (SD) is a measure of dispersion; that is, it provides information regarding how spread out the error scores were across all items in the sample. As the table shows, the SDs for all methods were quite large. This suggests that each method works well for some items, but quite poorly for other items. The current SAMMS method ranked sixth when SD is considered.

Given the large standard deviations observed, the average RMSE discussed above may not be a good measure by which to judge methods, since a few items with very large errors will have a large influence on this measure. Table 1 also shows the median scores for each method. The median (or 50th percentile) is that value which half the items in the sample score higher than; the other half obtain lower scores than the median. As the table shows, the decomposed 8-quarter MA performs quite poorly on this measure. The best methods are the two moving averages, SES, and the mean.

When all three measures (mean, SD, median) are considered together, single exponential smoothing appears to be the single "best" method. SES ranks in the top three on all three measures, and is the only method which does consistently well on all measures. The difference between the average RMSEs for the current SAMMS method and SES represents a 3.5% decrease in forecast error (recall that both methods are using optimized alpha values for each item).

Several additional points should be noted regarding the data

Table 3

ERRORS FOR 18
FORECASTING METHODS FOR SAMPLE 1

	RMSE			Rank Based on:	
Method	Mean	SD	Median	Median	SD
Dec MA8	114.9	629.6	7.71	16	1
SES	127.4	1885.0	6.93	3	2
MA4	129.6	1910.5	6.83	1	7
MA8	129.6	1910.5	6.83	2	8
SAMMS	132.0	1908.1	7.02	5	6
Gardner	134.3	1886.1	7.26	8	3
DES	135.5	1891.0	7.32	9	4
Trigg-Leach	137.9	1941.6	7.39	12	9
Dec Naive	139.8	2394.7	7.07	6	15
Mean	140.8	1907.6	6.98	4	5
Dec SES	142.3	2303.4	7.38	11	10
Dec MA4	143.6	2408.8	7.44	13	16
Dec SAMMS	144.4	2343.9	7.35	10	11
Dec Trigg-Leach	145.4	2373.9	7.50	14	13
Dec Mean	154.4	2361.0	7.60	15	12
Dec DES	164.7	2393.3	8.75	18	14
Holt	180.8	2598.2	8.53	17	17
Naive	274.8	7004.0	7.11	7	18

NOTE. Methods abbreviations are as follows:
SAMMS = current SAMMS method
SES = single exponential smoothing
DES = double exponential smoothing
MA4 = 4-quarter moving average
MA8 = 8-quarter moving average
"Dec" refers to decomposed or deseasonalized data.
See text for further explanation.

presented in Table 3. With the exception of the 8-quarter MA, all of the methods applied to the deseasonalized data performed poorly. The best of these methods, the naive forecast, ranked 9th on average RMSE and 6th on median RMSE. These results confirm the results of the previous analysis which showed that only a small proportion of the items in the sample have seasonal demand streams.

The above observation also holds true for trend as well as seasonality. The results of the autocorrelation analysis showed that less than 2% of the items in the sample have identifiable trends in their demand streams. Table 3 indicates that methods which forecast trend, including Holt's and Brown's double exponential smoothing, do quite poorly. By contrast, methods which ignore any trend in the data, including SES, the moving averages, and the current SAMMS method, are relatively accurate. This is further supported by the superiority of Gardner's model, which damps the trend term, over the Holt method.

One reason that the trend methods do poorly is that there are few items in the sample which exhibited any clear trend over the time period under study. A second reason for the poor performance of these methods is that the large amount of noise in the data tends to inflate the trend terms of those methods which employ them, thereby increasing the error for these methods.

3. Second Preliminary Analysis

The statistics provided in Table 3 represent one measure for comparing the forecast accuracy of the various methods. An alternative approach might be to compare the methods based on how often each method was the best one for each item. That is, how often did each method produce the smallest RMSE of all 18 methods? This question was answered by ranking the 18 methods from lowest to highest RMSE for each item separately, then counting the number of times a method received a rank of 1. The results of this procedure are shown in Table 4. The table shows, for each method, the number and percentage of items for which that method produced the most accurate forecast. Note that it is quite possible for two or more methods to be tied for "first place" (especially since error calculations were carried out to only two decimal places). The numbers shown in the table include the number of times a method tied for first place, regardless of the number of methods involved. For this reason, the numbers in the table do not sum to the total number of items in the sample. The methods are listed in order of decreasing first place scores.

The method which produced the best forecast for the largest number of items was the naive (last period's demand) method. This is perhaps rather surprising in view of the poor performance of this method when judged against the criteria presented in Table 3. The naive method's average RMSE and SD of errors is significantly higher than those of the other methods. The conclusion to be drawn from these two sets of data is that for

Table 4

NUMBER OF FIRST PLACE RANKS
FOR EACH OF 18 FORECASTING METHODS

Method	Number	Percent
Naive	981	10.0%
MA4	795	8.1%
MA8	795	8.1%
Dec Naive	772	7.9%
Holt	768	7.8%
Trigg-Leach	749	7.6%
Dec DES	653	6.7%
Mean	646	6.6%
Gardner	524	5.3%
DES	467	4.8%
Dec Mean	457	4.7%
Dec MA8	443	4.5%
Dec MA4	357	3.6%
Dec Trigg-Leach	346	3.5%
SAMMS	315	3.2%
SES	284	2.9%
Dec SES	279	2.8%
Dec SAMMS	180	1.8%

some items, those whose demand changes very little from one period to the next, the naive method produces better forecasts than any other method (note that items with many quarters of zero demands fit this category). For the remaining items, however, the naive method performs very poorly; the errors that it does make are large ones.

This same type of explanation can also be applied to the relatively poor performance of SES shown in Table 4. Given the findings shown in Table 3, this method appears to be a mediocre performer for all items. It does not do very well for very many items, but neither does it do very poorly for many items. In short, the data presented in the two tables clearly represent two different criteria for judging forecast accuracy.

Both of the moving average methods perform quite well when the data in Table 4 are considered. Further analysis of these two methods showed that for those items for which they are the most accurate, they are almost always tied with each other for this distinction. For example, the two methods tied for 1st place on 570 items, and each was the single best method for only eight items. This seems to clearly indicate that both moving averages are providing the same information in the forecast, and are equivalent in forecast accuracy.

As stated at the outset of this section, one of the goals of these preliminary analyses is to select a subset of "best" forecasting methods to be used in subsequent phases of the study. As the comparison of the two sets of results in Tables 3 and 4 suggests, however, the term "best" depends upon the particular application. If we were interested in implementing one single method for all items, then the criteria of Table 3 would be appropriate, and single exponential smoothing would probably be the method of choice. If, however, we were willing to maintain multiple forecasting methods, we could select the best method for a group of items by using the criterion of Table 4. Each method would be used to forecast only those items for which it was the most accurate. In this latter system, each method would be extremely inaccurate for some items, but that method would never be applied to those items. In short, the more forecasting methods we are willing to apply, the more likely we are to increase forecast accuracy, since some methods work best for some items, while other methods work best for other items.

4. Third Preliminary Analysis

The results presented in Table 4 do not directly address this issue, since they do not indicate how different forecasting methods perform in conjunction with each other for all items considered. Thus a third criterion for assessing the best subset of methods is to determine which group of methods, produces the lowest forecast error, when each method in the group is used to forecast only those items for which it is the most accurate method. One way to accomplish this would be to test all possible combinations of methods of a given size, and compare the RMSEs resulting from each combination. That combination with the smallest average RMSE is the best set of methods for that given size. For example, all possible combinations of two forecasting methods could be formed. For each item, the method which produced the smaller of the two RMSEs would be used to make the forecast. The average error across all items would then be computed. The procedure would be repeated for all remaining pairings of methods, and the pairing that produced the smallest RMSE would be the best possible combination of two methods. The entire procedure would then be repeated with all combinations of three forecasting methods, and so on. Since each additional method will forecast some items more accurately, adding an additional method will always decrease the forecast error. At some point, however, the cost of maintaining an additional forecasting method would outweigh the relative gain in forecast accuracy.

The procedure described above is the one that was used here to select the best subset of forecasting methods. An additional question, however, was whether to use the actual RMSE scores or the ranks in the procedure. The problem with using the actual error scores is that there may be some items which have extremely large errors for some methods, and these items may unduly influence the choice of the forecasting techniques. Using the ranks of the scores (when the methods are ranked within each item) avoids this problem, since the range of scores is the same for each item. Using ranks, however, means *losing a great deal* of valuable information concerning the magnitude of the differences between methods for each item.

This problem was resolved for purposes of the present analysis by standardizing the error scores within each item. Standard scores, or z-scores, measure how far each raw score is from the mean of the raw scores, in standard deviation units. The mean of z-scores is 0 and the standard deviation is 1.

The standardizing procedure was carried out for each item separately. This was accomplished by first calculating the mean and the standard deviation of the 18 errors generated by the forecasts. Z-scores were then computed by subtracting the mean from each error score and dividing this difference by the standard deviation. The resulting score indicates how far from the mean the error score is in standard deviation units. For example, if the mean of the error scores was 50 and the standard deviation was 10, then an error score of 60 would have a z-score of +1, while an error score of 45 would have a z-score of -.50.

The z-scores were substituted for the raw error scores in the subsequent analysis. Using the standard scores rather than the raw scores decreases considerably the variation between items in the sample while maintaining much of the information regarding the magnitude of the differences between forecasting methods.

The procedure used here, then, tested all possible combinations of forecasting methods using standardized scores. All combinations of sizes one through seven were examined. The best combination of a particular size was the one which produced the smallest long-term RMSE. One remaining problem was how to address the fact that the two moving average methods using the deseasonalized data could not be applied to all items. For the purposes of this analysis, these methods were given a z-score of +5 for each item they could not forecast. The +5 score was chosen as representing a score 10% greater than the largest z-score obtained for any method which actually does forecast an item.

The results of this process are shown in Table 5. The table shows the best methods for subsets ranging from size 1 to 7. The third column of Table 5 provides the sum of the z-scores across all items in the sample. Since a negative z-score represents a RMSE score which is lower than the mean, the numbers in this column represent lower average RMSEs. The last column of the table shows the percent improvement that each subset of methods represents over the next smaller subset. For example, using the best subset of two methods (Mean and Dec SES) results in a 125% decrease in the sum of the 2-scores (across the entire sample) from using just the one best method (SES).

Table 5

STANDARD SCORES FOR SELECTION OF BEST SUBSET OF 18 FORECASTING METHODS

No. of Methods	Best Subset	Sum of Z-scores	Percent Change
1	SES	- 2094.7	-
2	Mean, Dec SES	- 4715.6	125.1%
3	Mean, Holt, Dec Naive	- 6014.2	27.5%
4	Mean, Holt, Naive, Dec Naive	- 6589.4	9.6%
5	Holt, Dec Naive, Naive, Dec Mean, SAMMS	- 7087.9	7.6%
6	Holt, Dec Naive, Mean, Naive, SAMMS, Dec DES	- 7452.4	5.1%
7	Naive, Dec SES, Dec Mean, SES, MA4, Mean, Trigg-Leach	- 7768.8	4.2%

Table 5 shows that SES is the best single method, thus confirming the conclusion previously drawn from the results presented in Table 3. Table 5 also shows the improvement in forecast accuracy which can be gained from the use of multiple forecasting methods. This can be seen more readily by comparing the raw RMSE scores, rather than the z-scores, for these methods. The average RMSE which results from the use of the best subset of four methods is 108.7, compared with the average RMSE of 127.4 for SES (see Table 3). This represents approximately a 15% reduction in forecast error.

Examination of the percentage changes shows that the relative decrease in forecast error slows considerably when subsets of more than three methods are examined. Going from three to four methods, however, means adding the naive method to the first three. Since the naive is the least costly method to compute, and since its inclusion results in an additional 9.6% reduction in the z-score total, it was decided to use the 4-method subset (mean, Holt's exponential smoothing, naive, and the naive with deseasonalized data) in subsequent analyses.

Comparison of this best subset of four methods with the results presented in Table 4 shows the two procedures to be in reasonably good agreement. The best possible subset contains three of the five top ranked methods shown in Table 3, as well as the 8th ranked method. These four methods together were the best or tied for the best method for about one-third of all the items in the sample. The only major inconsistency between the two sets of results is the failure of the moving average methods to be included in any of the best subsets until the one with seven methods. Table 4 shows that the moving average method produced the most accurate forecasts for 8% of the items in the sample.

The approach for finding the "best" subset suffers from at least one major drawback. The error figures of Table 5 might be considered the maximum possible error reduction achievable, given perfect knowledge of which method to use. The approach assumes the ability to determine perfectly which method should be used with each item. For purposes of the analysis presented here, it was possible to test all methods against all items. In actual practice, however, this cannot be done, and perfect classification is not possible. That is, the criteria used to determine which method to use for which items will not work perfectly. Some items will be forecasted using a method other than the best method.

5. Synopsis of Preliminary Analyses

Given the limitation noted above, it seems prudent to synthesize the findings of the three analyses presented in Tables 3-5 in order to choose a best subset of methods for subsequent phases of the study. Specifically, the results shown in Table 3 suggest that single exponential smoothing be included in this subset, as it is the best individual model for the sample overall. As noted previously, the results of the other two analyses agree reasonably well, and argue for the inclusion of the naive, mean, Holt, and deseasonalized naive methods. Additionally, 4-quarter moving average is the best method for a relatively large number of items. This suggests that this method be included in the best subset as well.

To summarize, the analyses reported here have compared several different methods for reducing the number of forecasting methods to be included in subsequent phases of the study. Based on these procedures, the following methods have been selected for inclusion:

- single exponential smoothing
- 4-quarter moving average
- naive (last period's demand)
- mean
- Holt's exponential smoothing
- naive using deseasonalized data

B. Results Of Averaging Forecast Methods

The next step in the analysis was to examine various combinations of the six forecasting methods, identified in the previous subsection, by averaging the forecasts produced by the methods. These methods, along with their abbreviations, are as follows:

- single exponential smoothing (SES)
- 4-quarter moving average (MA4)
- naive (last quarter's demand) (NAIVE)
- mean (MEAN)
- Holt's exponential smoothing (HOLT)
- naive using deseasonalized data (DECNAIVE)

Two different procedures for averaging the forecasts produced by the different methods were examined here. The first was a simple average of the forecasts (AVG); averages of two methods at a time and three methods at a time were examined. The other used a weighted average (WTDAVG) with the weights based on the error produced by the method during the previous period. The equation for the weight for each forecast was:

$$w_{t,i} = 1 - (e_{t-1,i} / \sum_{j=1}^{n} e_{t-1,j})$$

where

$w_{t,i}$ is the weight assigned to the forecast from method i for period t

$e_{t,i}$ is the error for method i in period t

n is the number of methods involved.

The statistical measure used to compare the forecast error between different methods was the root mean square error (RMSE).

The RMSEs for the unweighted and weighted averages of the combinations of two methods are shown in Table 6. In all cases, adding a third method to the average did not decrease the error significantly, and these results are not shown here.

The single best combination using both unweighted and weighted averages used single exponential smoothing and the 4-quarter moving average. The RMSE for the average of these two methods was just slightly lower than the error for single exponential smoothing itself. The unweighted and weighted averages represent improvements of 3.9% and 3.7%, respectively, over the current SAMMS method. Both the unweighted average (AVG) and the weighted average (WTDAVG) of single exponential smoothing and the 4-quarter moving average will be included in the attempt to predict item groupings, which is the next step in the analysis.

C. Prediction Of Item Groupings And Forecast Methods

The goal of this phase of the analysis was to examine the relationship between item characteristics and item groupings based on forecast methods. This was accomplished using discriminant analysis (DA).

DA is a multivariate data analysis technique which is used to discriminate (and predict) between two or more groups based on a set of independent variables. This procedure allows the user to select from among a large set of variables those which are useful in separating the groups. This is done by finding linear combinations of variables which yield similar values for items in the same group, and different values for items in different groups (27). These linear combinations, known as discriminant functions, can then be used to develop a set of classification rules which can be used to predict group membership. The comparison of the predicted group membership and actual group membership is a measure of the usefulness of the analysis.

The first step in the analyses performed here was to place each item into a group. The number of groups was determined by the number of forecasting methods compared in each analysis. For each item, the RMSEs for the different forecasts were examined, and the item was placed into the group corresponding to the forecasting technique which produced the smallest error for that item. For example, one analysis looked at two methods: the 4-quarter moving average (MA4) and single exponential smoothing (SES). Thus there were two groups of items, corresponding to these two methods. Each item was placed in one of the two groups depending on which of the two forecast methods produced the smaller error. One group consisted of all those items forecasted more accurately by SES, while the other consisted of items forecasted more accurately by MA4.

Table 6

RMSES FOR AVERAGES OF TWO FORECASTING METHODS

UNWEIGHTED AVERAGES

Method	SAMMS	SES	MA4	MEAN	NAIVE	HOLT	DECNAIVE
SAMMS	131.97						
SES	129.08	127.39					
MA4	129.35	126.76	129.56				
MEAN	132.51	131.76	131.16	140.85			
NAIVE	134.16	131.36	133.32	134.16	143.28		
HOLT	150.81	147.82	149.78	149.01	154.25	180.77	
DECNAIVE	131.55	129.23	130.61	134.18	134.75	150.88	139.78

WEIGHTED AVERAGES*

	SAMMS	SES	MA4	MEAN	NAIVE	HOLT
SAMMS	131.97					
SES	129.91	127.39				
MA4	127.63	127.12	129.56			
MEAN	163.11	163.05	159.26	140.85		
NAIVE	130.14	130.13	133.72	162.31	143.28	
HOLT	136.74	135.82	138.97	157.11	144.24	180.77

---

* The DECNAIVE method was not included in the weighted averages.

Note: Numbers on the diagonal are the mean RMSEs for that method by itself.

Once the actual group membership for each item was established, the DA procedure was utilized to select those item characteristics most useful in predicting which group the item should be in, and to develop discriminant functions which could then be used to predict group membership. The success of this prediction was measured by examining the percentage of items which were correctly classified into each group. In addition, forecast errors were calculated for each item using the predicted group to select the forecast method, and these errors were then compared.

A total of 15 variables were included in the discriminant analyses:

- months since system entry
- months since last demand
- proportion of quarterly demands + or - 3 standard deviations (SDs) from mean demand
- unit price
- weapon system item (yes-no)
- supply status code ('1' vs. anything else)
- VIP code (VIP vs. non-VIP)
- demand quantity (last 4 quarters)
- demand frequency (last 4 quarters)
- number of quarters of demand data available
- "coefficient of variation" (SD of demand - mean of demand)
- "first difference ratio" (mean of first differences - mean of demand)
- three variables based on the first four autocorrelation functions (ACFs):
    - seasonal (yes-no; fourth ACF significant)
    - trend (yes-no; 1st 3 ACFs significant)
    - other (yes-no; other ACF significant)

The DA procedure was used to select the best subset of these items to use in predicting group membership. Table 7 presents the results of these analyses.

The first two columns of the table show the best possible RMSE (total and average for all items) for each method -- that is, the error obtained when each item is forecasted with the one method in that grouping which produces the lowest error for that item. Note that these are the errors which would result from the ability to predict which forecast method to use for each item with 100% accuracy. As the entries at the bottom of the table indicate, the combinations of three, four and six methods had smaller best possible errors than any of the pairs of methods.

Column 3 shows the results of the DA in terms of the percentage of correct classifications - that is, how often, based on the best combination of item characteristics, the procedure placed the item into the correct forecast method/group. As these figures show, none of the forecast method groupings were

Table 7

FORECAST ERRORS USING
PREDICTED ITEM GROUPINGS

Grouping Method	Best Possible Error Sum (1)	Best Possible Error Mean (2)	% Correctly Classified (3)	Actual Error Sum (4)	Actual Error Mean (5)	% Worse than Best (6)	% Diff from SAMMS (7)
SES MA4 AVG	769301.1	119.98	56.2%	806752.1	125.85	4.9%	-4.6%
SES MA4	769301.1	119.98	56.2%	808775.1	126.13	5.1%	-4.4%
SES DECNAIVE	757968.1	118.21	57.9%	812513.7	126.72	7.2%	-4.0%
SAMMS DECNAIVE	762314.1	118.89	59.0%	812701.7	126.75	6.6%	-4.0%
SES MEAN	776948.4	121.17	54.1%	814297.6	127.00	4.8%	-3.9%
MA4 MEAN	751684.6	117.23	59.0%	814864.7	127.08	8.4%	-3.7%
SAMMS SES	793792.3	123.80	53.7%	815402.6	127.17	2.7%	-3.6%
SAMMS MEAN	762400.3	118.90	56.2%	815602.7	127.20	7.0%	-3.6%
MA4 DECNAIVE	775496.1	120.94	57.3%	820599.3	127.98	5.8%	-3.0%
SES NAIVE	759919.8	118.52	76.1%	828557.1	129.22	9.0%	-2.1%
MEAN DECNAIVE	759412.1	118.44	61.1%	831745.8	129.65	9.5%	-1.8%
NAIVE DECNAIVE	758577.9	118.31	57.3%	836209.5	130.38	10.2%	-1.2%
MEAN NAIVE	750634.6	117.07	61.1%	840030.3	131.01	11.9%	-0.7%
SAMMS MA4	782277.2	122.00	57.7%	841031.7	131.17	7.5%	-0.6%
MA4 NAIVE	771179.9	120.26	54.3%	847580.7	132.18	9.9%	0.2%
SAMMS NAIVE	791216.1	123.40	62.0%	854968.4	134.12	8.7%	1.5%
SES HOLT	768409.5	119.84	55.0%	888601.6	138.59	15.6%	5.0%
MEAN HOLT	755330.4	117.80	59.9%	909187.1	141.79	20.4%	7.4%
SAMMS HOLT	805955.8	125.70	58.4%	911036.1	142.08	13.0%	7.7%
MA4 HOLT	786679.8	122.69	53.7%	924610.2	144.20	17.5%	9.3%
NAIVE HOLT	818228.4	127.61	56.9%	950950.8	148.31	16.2%	12.4%
HOLT DECNAIVE	820294.6	127.93	54.7%	1033854.1	161.24	26.0%	22.2%
MEAN HOLT DECNAIVE	718588.7	112.07	41.9%	873764.8	136.27	21.6%	3.3%
MEAN NAIVE HOLT DECNAIVE	692161.7	107.95	32.3%	864590.1	134.84	24.9%	2.2%
SES NAIVE MEAN MA4 HOLT DECNAIVE	679843.3	105.87	22.6%	823571.1	128.45	21.1%	-2.7%

Note: Numbers in parentheses are column numbers. See text for references.

predicted very accurately by the DA procedure. Generally speaking, classification accuracy decreased as the number of groups increased. This is due to the fact that the groups are not very different from each other, at least not in ways which can be predicted from the item characteristics available.

The next two columns of the table show the actual error that was obtained by using the classifications produced by the DA. The entries in the top part of the table (that is, the 2-method groupings) are listed in order of increasing actual error. The best method was a combination of single exponential smoothing (SES), the 4-quarter moving average (MA4), and the average of these two methods. This particular method (SES/MA4/AVG), which was developed in conjunction with the DA procedure, will be described in detail shortly.

None of the 3, 4, or 6-method subsets performed particularly well. The best of these (all 6 methods) ranked tenth overall in actual error. The reason for this is shown in the next column of the table. This column shows the percent difference between the best possible error (given perfect prediction) and the actual error. The groupings of 3, 4 and 6-methods had at least 20% difference between the best possible error and the actual error.

Finally, the last column of the table shows the percentage difference between each grouping of methods and the current SAMMS method. Note that in the comparisons with SAMMS in this table (and all subsequent tables as well), negative signs indicate improvement in forecast accuracy. For example, the combination of SES, MA4, and the average of the two produced an error which was 4.6% lower (more accurate) than the SAMMS method.

This best method, the combination of SES, MA4 and their average, requires some explanation. The SES/MA4 average was the best of all averaging methods. Since the accuracy of the classification procedure resulting from the DA was low, an alternative forecasting method was developed. This method used SES or MA4 to forecast the item only if the probability of making the classification was reasonably high. If the choice between the methods could not be made with confidence, then the average of the two forecasts was used.

The probabilities referred to above were obtained from the DA procedure. The analysis generated a linear prediction equation for each group known as a classification equation. A classification score was then computed for each group by multiplying an item's values for the variables in the equation by the corresponding coefficients. Given certain statistical assumptions, each classification score can be converted into the probability that an item belongs in a group.

The procedure employed here involved calculating these probability values for the item groupings corresponding to the two forecast methods, SES and MA4. The cutoff values used to decide whether or not to classify were determined empirically. That is, all possible combinations of the two probability values were examined, and the errors compared. The two probabilities which produced the lowest error were used as cutoff points. This procedure resulted in decision rules as follows:

- if the probability of being in the SES group was equal to or greater than .55, use SES to forecast.
- if the probability of being in the MA4 group was equal to or greater than .75, use MA4 to forecast.
- if neither of the above, use the average of SES and MA4 to forecast.

These decision criteria resulted in the selection of the average for the majority (62%) of the items. Exponential smoothing was used for 36% of the items, while the MA4 method was selected for only 2% of the items in the sample.

The results shown in Table 7 suggest that the use of multiple (i.e., more than two) forecasting methods could potentially produce significantly more accurate forecasts than the use of a single method. The variables used here, however, could not successfully predict which method to use for which item. The actual errors observed, therefore, suggest the use of no more than two forecast methods at a time.

The DA procedure described above will lead to overly optimistic results, since all classification scores are optimal for the items in this particular sample. In fact, all of the analyses presented to this point will be biased, since all procedures were developed and then tested on the same sample. Therefore, it was necessary to evaluate the methods using the additional samples selected randomly from the population.

D. Validation Of Findings

1. Results for Additional Samples

Based on the results presented in the previous sections, the following forecast methods were selected to be tested on subsequent samples:

- SAMMS method (alpha = .2)
- SAMMS method (alpha = .1)
- SAMMS method (individual alpha for each item)
- SES (alpha = .1)
- SES (individual alpha for each item)
- MA4
- AVG (alpha = .1)
- AVG (individual alpha for each item)
- WTDAVG (alpha = .1)
- WTDAVG (individual alpha for each item)
- SES/MA4/AVG (alpha = .1)
- SES/MA4/AVG (individual alpha for each item)

- SES/MA4/WTDAVG (alpha =.1)
- SES/MA4/WTDAVG (individual alpha for each item)

The "alpha" referred to above is the smoothing constant used in the equations for the corresponding techniques. Note that these methods represent the current system (SAMMS), two alternative methods (SES and MA4), weighted and unweighted averages of the alternative methods, and the combination of the methods based on the classification function developed using the items from the first sample. In addition, each of these was examined (1) using the best single value for the smoothing constant alpha, and (2) using individual smoothing values for each item.

As discussed previously, two additional random samples (of 6,815 and 6,499 items, respectively) were drawn from the population. Each of the forecast methods named above was applied to the items in the second and third samples. The resulting forecasts for the last four periods were compared to the actual demand, and the four step-ahead RMSEs were calculated. In the case of the SES/MA4/AVG method, the cutoff probabilities developed from the first sample were applied to the other two samples.

The RMSEs for the original sample, along with the two additional samples, are shown in Table 8. Examination of the average errors shows that the magnitude of the forecast error varied considerably from sample to sample. Specifically, the error was greatest in the second sample, and smallest in the third sample. In addition, the relative error of the various alternative methods compared to the SAMMs baseline method varied across the three samples. For example, the maximum improvement of any method over the current SAMMS procedure was 3.2%, 4.2%, and 1.1%, respectively, in the three samples.

The table also shows the rankings of the various forecast methods in each sample. Rankings are from the smallest RMSE (with a rank of 1) to the largest RMSE (with a rank of 13; the baseline method was not ranked). These numbers show that the relative performance of the forecasting methods also varied across the three samples. For example, the SES/MA4/WTDAVG method, which produced the smallest forecast error in the first sample, was ranked 3 and 7 in the second and third samples, respectively.

Despite these types of differences, similarities across the samples are also apparent. The best overall methods were the weighted average, using a smoothing constant of .1 or individual alphas for each item. Other consistently good performers were the SES/MA4/WTDAVG method, again using both the .1 and individual alphas, and the SES/MA4/AVG method using individual alphas. The best of the single methods was the SES method with the individual alphas for each item. Several methods were also consistently poor performers, as indicated by the rankings in Table 3. These included the SAMMS double exponential smoothing, with individual alphas and an alpha of .1, SES with an alpha of .1, and the 4-quarter moving average.

Table 8

RMSES FOR THE THREE SAMPLES

METHOD	SAMPLE 1 MEAN	SAMPLE 1 SUM	SAMPLE 1 % DIFF FROM SAMMS	SAMPLE 1 RANK	SAMPLE 2 MEAN	SAMPLE 2 SUM	SAMPLE 2 % DIFF FROM SAMMS	SAMPLE 2 RANK	SAMPLE 3 MEAN	SAMPLE 3 SUM	SAMPLE 3 % DIFF FROM SAMMS	SAMPLE 3 RANK
SAMMS BASELINE	129.70	831648.33	-	-	145.89	930184.13	-	-	119.15	729565.44	-	-
SAMMS(.1)	130.02	833669.02	0.24%	11	146.49	934016.52	0.41%	12	119.90	734134.86	0.63%	12
SAMMS(I)	131.97	846200.35	1.75%	12	145.25	926134.62	-0.44%	10	119.72	733016.42	0.47%	11
SES .1	132.94	852431.78	2.50%	13	145.50	927700.42	-0.27%	11	121.39	743281.02	1.88%	13
SES(I)	127.40	816871.38	-1.78%	7	139.83	891576.73	-4.15%	2	118.85	727718.04	-0.25%	9
MA4	129.56	830761.45	-0.11%	10	146.95	936925.71	0.72%	13	119.42	731209.77	0.23%	10
AVG(.1)	128.90	826530.33	-0.62%	9	143.50	914959.27	-1.64%	8	118.09	723068.59	-0.89%	3
WTDAVG(.1)	126.79	812950.08	-2.25%	6	139.76	891135.80	-4.20%	1	118.11	723203.83	-0.87%	4
AVG(I)	126.76	812796.28	-2.27%	5	141.78	904004.87	-2.81%	7	117.95	722226.55	-1.01%	2
WTDAVG(I)	126.28	810357.12	-2.56%	4	140.42	895291.74	-3.75%	4	117.85	721620.94	-1.09%	1
SES/MA4/AVG(.1)	128.01	820777.50	-1.31%	8	143.54	915180.37	-1.61%	9	118.80	727431.88	-0.29%	8
SES/MA4/AVG(I)	125.93	807459.00	-2.91%	2	140.91	898400.54	-3.42%	5	118.38	724868.16	-0.64%	5
SES/MA4/WTDAVG(.1)	125.95	807615.68	-2.89%	3	140.96	898753.63	-3.38%	6	118.46	725340.54	-0.58%	6
SES/MA4/WTDAVG(I)	125.57	805156.00	-3.19%	1	140.12	893423.14	-3.95%	3	118.52	725702.41	-0.53%	7

Note: Values in ( ) are alpha values. "I" indicates the use of individual alphas for each item.

One additional point of interest in this analysis is the effect of using individual smoothing parameters for each item, versus using a single parameter for all items. Aside from the current SAMMS method, there are five methods which can be used to assess this effect: SES, AVG, WTDAVG, SES/MA4/AVG, and SES/MA4/WTDAVG. Table 9 presents the percentage changes resulting from using individual versus fixed alphas for each of these five forecast methods.

Table 9

COMPARISON OF SINGLE VS. INDIVIDUAL
ALPHA VALUES FOR THE THREE SAMPLES

Method	Sample 1	Sample 2	Sample 3	Average
SES	-4.2%	-3.9%	-2.1%	-3.4%
AVG	-1.7%	-1.2%	-0.1%	-1.0%
WTDAVG	-0.3%	0.5%	-0.2%	-0.1%
SES/MA4/AVG	-1.6%	-1.8%	-0.3%	-1.3%
SES/MA4/WTDAVG	-0.3%	0.6%	0.1%	-0.3%
Average	-1.6%	-1.4%	-0.5%	-1.2%

Note. Entries are percent changes resulting from the use of individual alphas versus a single alpha value for all items. Negative percentages indicate more accurate forecasts using individual alphas.

The only method for which the use of individual alpha values makes a significant difference is SES. The average percentage improvement across the three samples was 3.4% for this method. While the use of individual alphas lowered the RMSE in almost all cases, the improvement was not very large. As the last row in the table shows, the average improvement gained by using individual alphas was 1.2%, with a large proportion of this improvement being due to the exponential smoothing method. Without SES, the improvement due to using individual alphas for each item is 0.7%.

The explanation for the finding that individual alphas do not improve forecast accuracy any more than they do relates to the manner by which the alphas were selected. In preliminary runs alpha values from .1 to 1, in increments of .1, were examined, and the forecast error over all periods except for the last four was calculated. The RMSEs used to compare methods here, however, are calculated over the last four periods only. Therefore, if the last four periods do not maintain the same pattern as the previous periods, the "best" alpha for the two time intervals willnot necessarily be the same.

Some evidence for this hypothesis comes from the analysis of the autocorrelation functions (ACFs) for the items in the samples. As noted in subsection A, 79% of the items in Sample 1 failed to show any significant ACFs, indicating a random pattern to the demand data for these items. Repeating this analysis for Samples 2 and 3 showed random demand patterns for 79.2% and 79.7% of the samples, respectively. Since the data streams for the vast majority of items are random, there is no reason to expect the last four data points would look like the initial points.

Although Table 8 did show some similarities in findings from one sample to the next, there were also enough differences to be of concern. These inter-sample differences were believed to be due to the randomness in the demand data for the items in the population as a whole, rather than to any problems with the sampling process. Given the randomness of the data, as discussed previously, it was difficult to determine which, if any, of the three samples' results were representative of the entire population. Since there were differences between samples as well, it seemed prudent to test the methods shown in Table 8 using the entire population of items to be forecasted by DLA. This analysis is presented in the next section.

2. Results for Population

Since the population consists of over 677,000 items, it was not feasible to examine all of the methods shown in Table 8. Specifically, the identification of individual smoothing constants for each item is extremely time consuming and costly. It was, therefore, decided to test only methods which used a single alpha value for all items. Eliminating these left eight forecasting methods which were computed for all items in the population: SAMMS (.2 alpha), SAMMS (.1 alpha), SES (.1 alpha), MA4, AVG, WTDAVG, SES/MA4/AVG, and SES/MA4/WTDAVG. The latter two methods again employed the coefficients and cutoff scores derived from the discriminant analysis on the items in Sample 1.

The original population consisted of 677,705 items. Of these, 41,649 items (6.1%) were eliminated from the analysis either because they had only four quarters of demand, or because all quarters except for the last four had zero demand. This left 636,056 items for analysis.

The results of these analyses for the entire population of items are shown in Table 10. The results are reported in terms of both RMSE and the mean absolute deviation (MAD) of the forecast errors (the average over the last four periods of the absolute values of the differences between the forecast and the actual demand).

Table 10

FORECAST ERRORS OF SELECTED
METHODS FOR ENTIRE POPULATION

METHOD	RMSE SUM	MEAN	% DIFF FROM SAMMS	MAD SUM	MEAN	% DIFF FROM SAMMS
SAMMS(.2)	77488274.29	121.83	-	129328480.61	203.33	-
WTDAVG	75224054.51	118.27	-2.92%	124279128.72	195.39	-3.90%
SES/MA4/WTDAVG*	76759232.78	120.68	-0.94%	127361526.05	200.24	-1.52%
MA4	77073457.66	121.17	-0.54%	127701415.93	200.77	-1.26%
AVG	77341887.62	121.60	-0.19%	128734602.55	202.40	-0.46%
SES/MA4/AVG*	78274588.69	123.06	1.01%	130509540.38	205.19	0.91%
SAMMS(.1)	79982598.47	125.75	3.22%	134193435.75	210.98	3.76%
SES	81659832.91	128.38	5.38%	136997014.96	215.39	5.93%

----

* The number of items for which the methods were selected were as follows:
SES: 226,646 (35.6%); MA4: 10,651 (1.7%); Average: 398,759 (62.7%).

The single best method for the entire population was the weighted average of the forecasts from the SES and MA4 methods. This method produced a 2.9% lower RMSE than the baseline SAMMS method, and a 3.9% lower MAD than the SAMMS method. The next best method, the SES/MA4/WTDAVG, was clearly inferior to the WTDAVG procedure, as was the MA4 method by itself. All three of these methods, plus the unweighted average, produced smaller RMSEs and MADs than the current SAMMS method. By contrast, single exponential smoothing by itself was a poor performer, as was the current SAMMS method with a smaller alpha value.

The finding that the weighted average is the best method is consistent with the conclusions reached from the examination of the results for the three samples shown in Table 8. The weighted average was the best method in the second sample, and was ranked 6 and 4 in the first and third samples, respectively. These ranks made this method one of the most consistently effective across the three samples.

Although individual alphas for each item were not examined for all items in the population, the errors presented in Table 9 give some indication of how the WTDAVG might perform using individual alphas for SES for each item, rather than a single alpha (.1) for all items. As Table 9 shows, using individual alphas for the WTDAVG resulted in forecast error improvements of 0.3% in the first sample and 0.2% in the third sample. In the second sample, the individual alphas actually produced a 0.5% larger forecast error than the use of an alpha of .1 for all items. The average reduction in error across the three samples was 0.02%; for the first and third samples only, the average reduction was 0.25%.

The results for the entire population of items were examined further with regard to two key variables: commodity and weapon system. Tables 11 and 12 present the RMSEs and MADs for five forecast methods by commodity and weapon system status, respectively.

As Table 11 shows, the relative rankings of the various methods is consistent across all commodities with the exception of Medical. For the other five commodities, the WTDAVG and the SES/MA4/WTDAVG are the most accurate forecast methods. For the Medical commodity, the MA4 method by itself produced the lowest forecast error.

Although the rankings of the methods are similar across commodities, the ability of the methods to improve upon the current SAMMS forecasts varied considerably from one commodity to the next. This is shown in the two columns which list the percentage difference between each method and the current SAMMS method (note that both the RMSE and the MAD are absolute error measures; thus, the relative magnitude of the errors across commodities simply reflects differences in demand rates). All commodities improved to some extent, with the exception of the

Table 11

FORECAST ERRORS BY COMMODITY FOR POPULATION

METHOD	RMSE SUM	MEAN	% DIFF FROM SAMMS	MAD SUM	MEAN	% DIFF FROM SAMMS
			CONSTRUCTION (N=91,207)			
SAMMS	3895556.42	42.60	-	6544608.81	71.76	-
WTDAVG	3858049.88	42.30	-0.71%	6470385.36	70.94	-1.13%
SES/MA4/WTDAVG	3850007.87	42.21	-0.91%	6463844.19	70.87	-1.23%
MA4	3900871.39	42.77	0.39%	6527863.44	71.57	-0.26%
SES	4169945.82	45.72	7.32%	7085541.71	77.69	8.27%
			ELECTRONICS (N=168,633)			
SAMMS	4726760.09	28.03	-	7794730.72	46.23	-
WTDAVG	4692114.94	27.83	-0.73%	7671755.51	45.50	-1.59%
SES/MA4/WTD	4696311.77	27.85	-0.64%	7696314.01	45.64	-1.26%
MA4	4805868.42	28.50	1.67%	7876395.05	46.71	1.05%
SES	4924265.63	29.20	4.18%	8143456.49	48.29	4.47%
			GENERAL (N=79,266)			
SAMMS	7665376.53	96.70	-	12866650.63	162.32	-
WTD AVG	7271217.69	91.73	-5.14%	12090391.55	152.53	-6.03%
SES/MA4/WTD	7274441.34	91.77	-5.10%	12086701.39	152.48	-6.06%
MA4	7326630.74	92.43	-4.42%	12182769.45	153.69	-5.32%
SES	8476994.12	106.94	10.59%	14487873.65	182.78	12.60%
			INDUSTRIAL (N=276,654)			
SAMMS	52177306.75	198.60	-	86956868.17	314.32	-
WTD AVG	50364361.54	182.05	-3.47%	82887155.01	299.61	-4.68%
SES/MA4/WTD	51835800.33	187.37	-0.65%	85776314.16	310.12	-1.33%
MA4	52064550.47	188.19	-0.22%	86075664.05	311.13	-1.01%
SES	54276811.25	196.19	4.02%	90527029.07	327.22	4.11%
			MEDICAL (N=11,612)			
SAMMS	2922689.36	251.70	-	5044308.66	434.40	-
WTD AVG	3008089.63	259.05	2.92%	5230051.47	450.40	3.68%
SES/MA4/WTD	3066953.87	264.12	4.94%	5371088.93	462.55	6.49%
MA4	2915518.36	251.08	-0.25%	5019030.45	432.23	-0.50%
SES	3605612.81	310.51	23.37%	6450743.51	555.52	27.85%
			CLOTHING & TEXTILES (N=8,694)			
SAMMS	6110585.05	702.85	-	10121313.62	1164.17	-
WTD AVG	6030220.82	693.61	-1.32%	9979389.83	1147.85	-1.40%
SES/MA4/WTD	6035717.61	694.24	-1.23%	9947263.37	1144.15	-1.72%
MA4	6060018.28	697.03	-0.83%	10019693.79	1152.48	-1.00%
SES	6206203.29	713.85	1.56%	10302370.53	1185.00	1.79%

Table 12

FORECAST ERRORS BY WEAPON
SYSTEM STATUS FOR POPULATION

METHOD	RMSE SUM	MEAN	% DIFF FROM SAMMS	MAD SUM	MEAN	% DIFF FROM SAMMS
WEAPON SYSTEM (N=336,722)						
SAMMS	51284025.76	152.30	-	85640200.69	254.34	-
WTDAVG	50062696.75	148.68	-2.38%	82875682.81	246.12	-3.23%
SES/MA4/WTD	50898600.77	151.16	-0.75%	84594433.16	251.23	-1.22%
MA4	51241786.39	152.18	-0.08%	85064686.27	252.63	-0.67%
SES	54076907.38	160.60	5.45%	90844079.91	269.79	6.08%
NON-WEAPON SYSTEM (N=299,334)						
SAMMS	26204248.54	87.54	-	43688279.92	145.95	-
WTD AVG	25161357.75	84.06	-3.98%	41403445.92	138.32	-5.23%
SES/MA4/WTD	25860632.01	86.39	-1.31%	42767092.89	142.87	-2.11%
MA4	25831671.27	86.30	-1.42%	42636729.66	142.44	-2.41%
SES	27582925.53	92.15	5.26%	46152935.05	154.19	5.64%

medical commodity. The largest improvement in both RMSE and MAD was seen for the General commodity (5.1% and 6.0% for the RMSE and MAD, respectively). A 3.5% improvement in the RMSE was observed for the Industrial commodity. The other three commodities showed only slight improvement over the current SAMMS method. The WTDAVG and the SES/MA4/WTDAVG performed worse than the SAMMS method for the items in the Medical commodity. Only the MA4 method produced a lower forecast method than SAMMS, and this difference was very slight.

Table 12 presents the RMSEs and MADs for weapon system versus non-weapon system items. All methods, with the exception of SES, improved forecast accuracy over the current SAMMS method for both weapon and non-weapon system items. Once again, the WTDAVG method produced the greatest decrease in forecast error (for both types of items). The improvement over the current SAMMS forecast accuracy was greater for non-weapon system items (4.0%) than it was for weapon system items (2.4%)

To summarize, a total of 14 forecast procedures were examined for three separate random samples of items. Eight of these, using fixed alpha values for all items, were examined for the entire population of 636,056 items. The results of the latter analysis showed that the weighted average of the forecasts from simple exponential smoothing (alpha = .1) and the 4-quarter moving average produced the smallest error, as measured by both the root mean square error (RMSE) and the mean absolute deviation of forecast errors (MAD). This method produced a RMSE which was 2.9% smaller than that produced by the current forecast method used in SAMMS, and a MAD that was 3.9% smaller than the SAMMS baseline method. Based on findings from the three samples, it is unlikely that the use of individual smoothing constants for each item would improve the forecast accuracy of this method by more than 0.25%. A breakdown of these results by commodity showed that the greatest improvement over the current SAMMS method was seen for the General and Industrial commodities. The WTDAVG method performed more poorly for the Medical commodity than the current SAMMS method.

E. Impacts Of Forecast Methods On Inventory System

The findings reported in the previous section are based on statistical criteria, such as the RMSE and the MAD. While such measures are important, they are not the only ones which must be considered in an inventory system. The ultimate goal of improving forecasting in DLA is to improve customer service, or to maintain customer service at a satisfactory level while reducing the costs of the service. In inventory terms, this translates into increasing supply availability and decreasing backorders, or holding supply availability constant and reducing safety level stocks. These types of variables are as important in evaluating the value of a forecasting technique as the statistical accuracy measures already presented.

The impact of an overall decrease in the MAD on safety level requirements can be assessed in a preliminary way using basic inventory equations. Assuming leadtime demand is normally distributed with mean μ and standard deviation σ, the equation for the reorder point, r, is

$$r = \mu + z\sigma ,$$

where z is the number of standard deviations necessary to achieve a desired level of customer support. The second term in the equation, $z\sigma$, represents the safety level. In the SAMMS system the MAD is a reasonable estimate of σ. Therefore, for a constant customer support level, any reduction in the MAD would be expected to produce a proportional reduction in the safety level. In this case, the 3.9% reduction in MAD which would be obtained from substituting the WTDAVG for the current SAMMS forecast procedure should produce a corresponding 3.9% decrease in safety level.

The results of a recent empirical study (28) of the relationship betweeen safety level and MAD in the SAMMS system suggest larger decreases in safety level associated with lowering the MAD than those noted above. Using a sample of items from the hardware commodities, the study showed that each 5% reduction in MAD leadtime results in a 7.3% reduction in safety level dollars (28, Table 14). Using this as a guide, the 3.9% reduction in the MAD observed here for the weighted average method would result in a 5.7% decrease in safety level dollars. This assumes that leadtimes and alpha values (the other factors, aside from the MAD, that determine MADLT) would remain constant, and that the sample used in the analysis was in fact representative of the larger population.

Table 13 uses one of these estimates of safety level reduction to translate the observed changes in MADs for the various forecasting methods into safety level dollar changes. Total safety level dollars was calculated for each commodity by multiplying each item's safety level quantity by the item's standard price, and summing the results across all items in the commodity. The database files developed for the study were used, so that the prices and quantities were those in effect in the last quarter of 1984. The table assumes the proportional change in safety level dollars associated with the basic inventory equations discussed previously. It should be noted that these are conservative estimates when compared with the empirical results discussed above.

Table 13 shows that the General commodity has the largest estimated safety level savings, at over $12 million. The WTDAVG method produces consistently large savings in safety level dollars for all commodities except Medical. For Medical, the four-quarter moving average produces an estimated $607,000 savings in safety level. Overall, these figures suggest that substituting the MA4 method for the SAMMS method in the Medical

Table 13

ESTIMATED SAFETY LEVEL CHANGES BY COMMODITY

METHOD	% DIFF FROM SAMMS	ESTIMATED SL CHANGE
	CONSTRUCTION	
TOTAL SAFETY LEVEL	-	$239,477,301
WTDAVG	-1.13%	($2,715,950)
SES/MA4/WTDAVG	-1.23%	($2,955,302)
MA4	-0.26%	($612,739)
SES	8.27%	$19,793,567
	ELECTRONICS	
TOTAL SAFETY LEVEL	-	$121,241,768
WTDAVG	-1.58%	($1,912,796)
SES/MA4/WTD	-1.26%	($1,530,803)
MA4	1.05%	$1,270,233
SES	4.47%	$5,424,194
	GENERAL	
TOTAL SAFETY LEVEL	-	$210,745,219
WTDAVG	-6.03%	($12,714,489)
SES/MA4/WTD	-6.06%	($12,774,931)
MA4	-5.32%	($11,201,415)
SES	12.60%	$26,554,308
	INDUSTRIAL	
TOTAL SAFETY LEVEL	-	$125,763,546
WTDAVG	-4.68%	($5,885,924)
SES/MA4/WTD	-1.33%	($1,678,480)
MA4	-1.01%	($1,274,464)
SES	4.11%	$5,163,435
	MEDICAL	
TOTAL SAFETY LEVEL	-	$121,223,421
WTDAVG	3.68%	$4,463,719
SES/MA4/WTD	6.48%	$7,853,092
MA4	-0.50%	($607,486)
SES	27.88%	$33,799,051
	CLOTHING & TEXTILES	
TOTAL SAFETY LEVEL	-	$113,158,680
WTD AVG	-1.40%	($1,586,742)
SES/MA4/WTD	-1.72%	($1,945,923)
MA4	-1.00%	($1,136,134)
SES	1.79%	$2,024,259

Note: Parentheses indicate cost savings.

commodity and the WTDAVG method for the SAMMS method in the other commodities would result in an estimated savings in safety level of $25,423,388.

In order to examine these types of inventory variables in greater detail, the present study made use of a simulation model. The Uniform SAMMS Inventory Management Simulation (USIMS) is a simulation model which can be used to examine the impacts of alternative inventory policies (in this case forecast methods) on the performance of the various Defense Supply Centers (DSCs). The model uses a sample of DLA's items in conjunction with a Monte Carlo simulation of key inventory events (29, p. 1). The sample is a stratified random sample of items, with the stratification based on annual dollar demand.

For the purposes of this study, the basic USIMS model was altered in several ways. First, the various alternative forecasting methods discussed in the previous section were added to the model. In addition, actual demand data was substituted for the stochastically generated data normally used in the model. To accomplish this, all requisitions for the items in the USIMS sample were obtained for the two-year period from July, 1983 through June, 1985. These actual requisitions served as the input data for the model.

Before presenting the results of the simulation analysis, it should be noted that the findings here must be interpreted with caution. USIMS, like any other simulation, is an imperfect model of the "real world" system. Moreover, the analysis performed here necessitated making various assumptions about how a new forecasting system would be implemented in SAMMS. These assumptions, which will be discussed later, were not subjected to a rigorous testing process, and may, therefore, be invalid.

The purpose of the simulation analysis was to obtain an impression of the relative impacts of the various forecasting methods on the inventory system. The figures resulting from the analysis should only be used to compare methods with each other. There is no guarantee that the magnitude of the differences reported here would actually be realized should a particular forecasting method be implemented.

Figures 1 thru 5 present the results of the simulation analysis for five key variables: supply availability, safety level dollars, total dollar value of commitments, number of backorders, and average number of days on backorder. Each graph depicts the performance of four methods: the SAMMS baseline, single exponential smoothing, the four-quarter moving average, and the weighted average of the latter two methods. The numbers which are graphed in these figures were obtained by averaging the values of the relevant variable across the eight quarters of the simulation. These averages were then totaled across all of the commodities for safety level dollars, commitments, and number of backorders, and averaged across the commodities for supply

FIGURE 1

AVERAGE SUPPLY AVAILABILITY

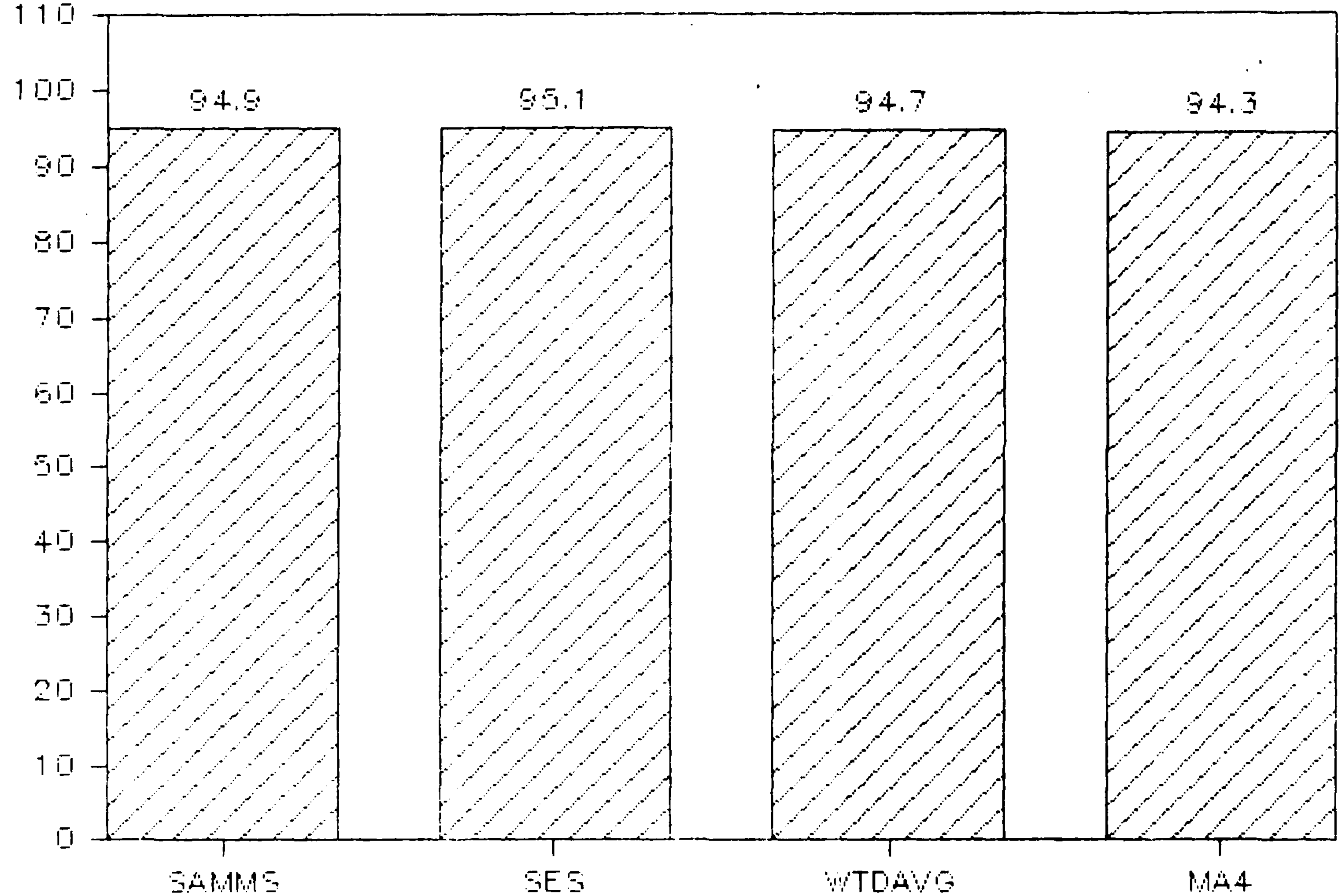


FIGURE 2

AVERAGE SAFETY LEVEL DOLLARS

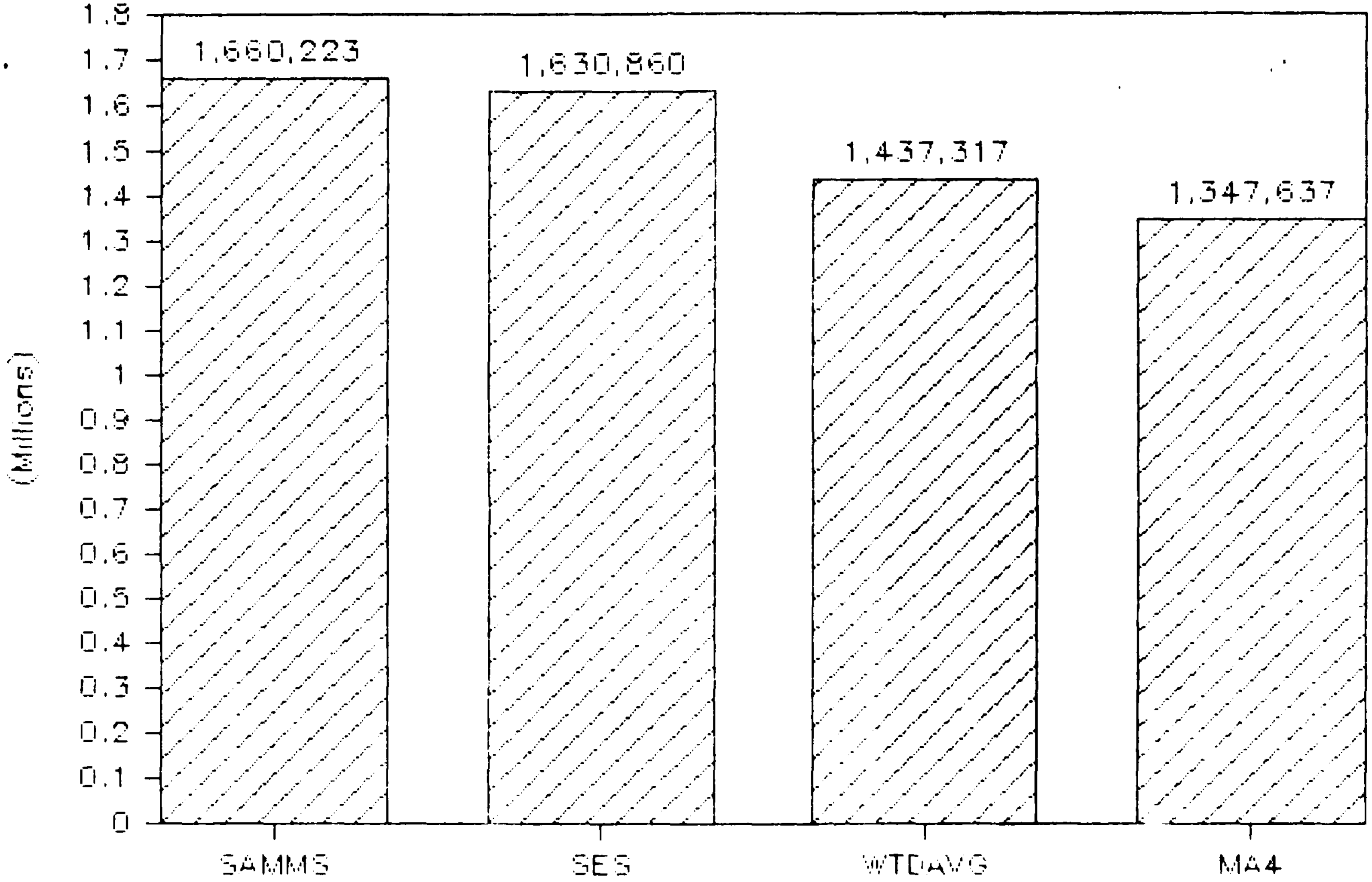


FIGURE 3

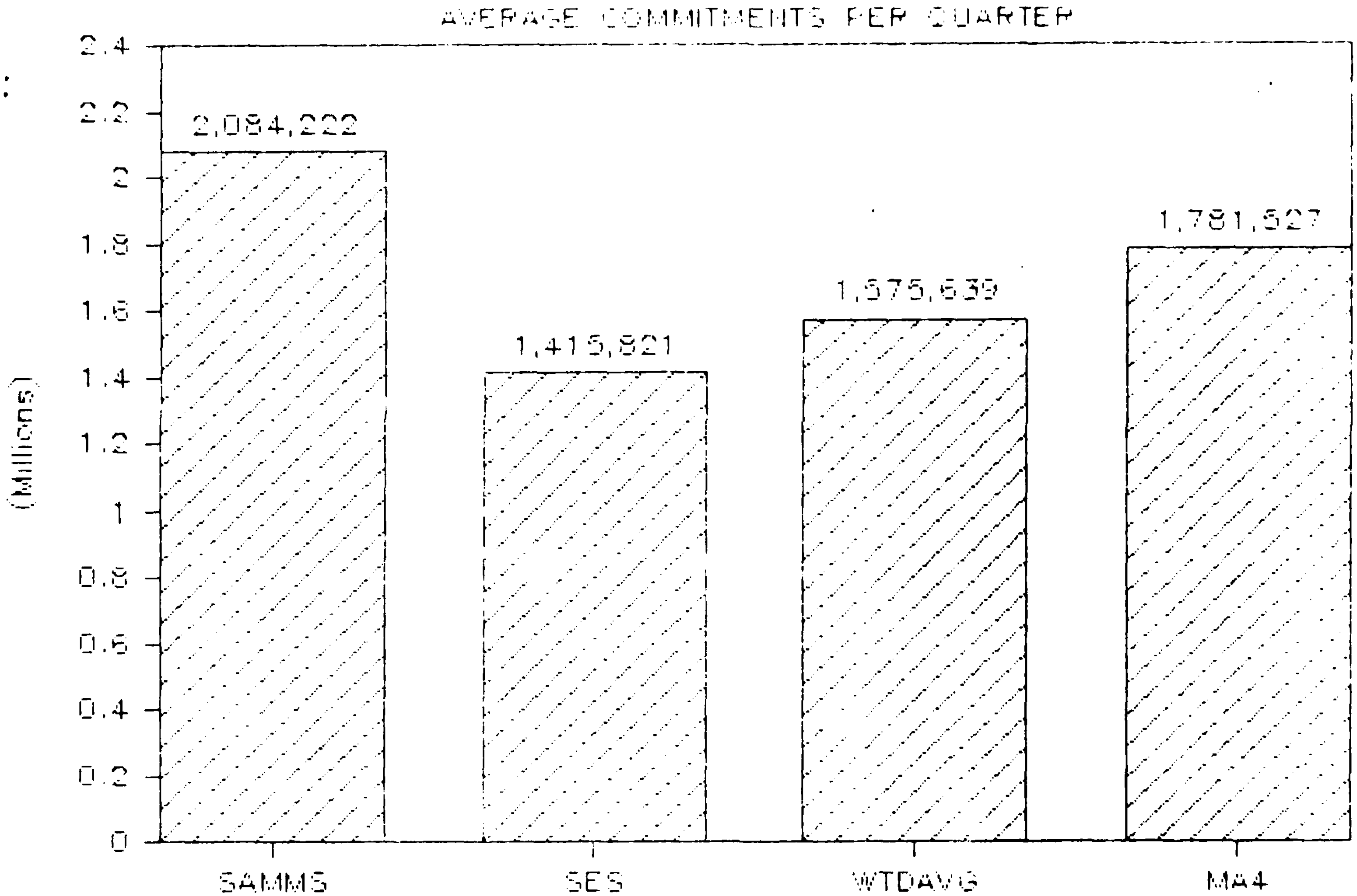

AVERAGE COMMITMENTS PER QUARTER
(Millions)
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
2,084,222
1,415,821
1,575,639
1,781,527
SAMMS
SES
WTDAVG
MA4

Figure 4

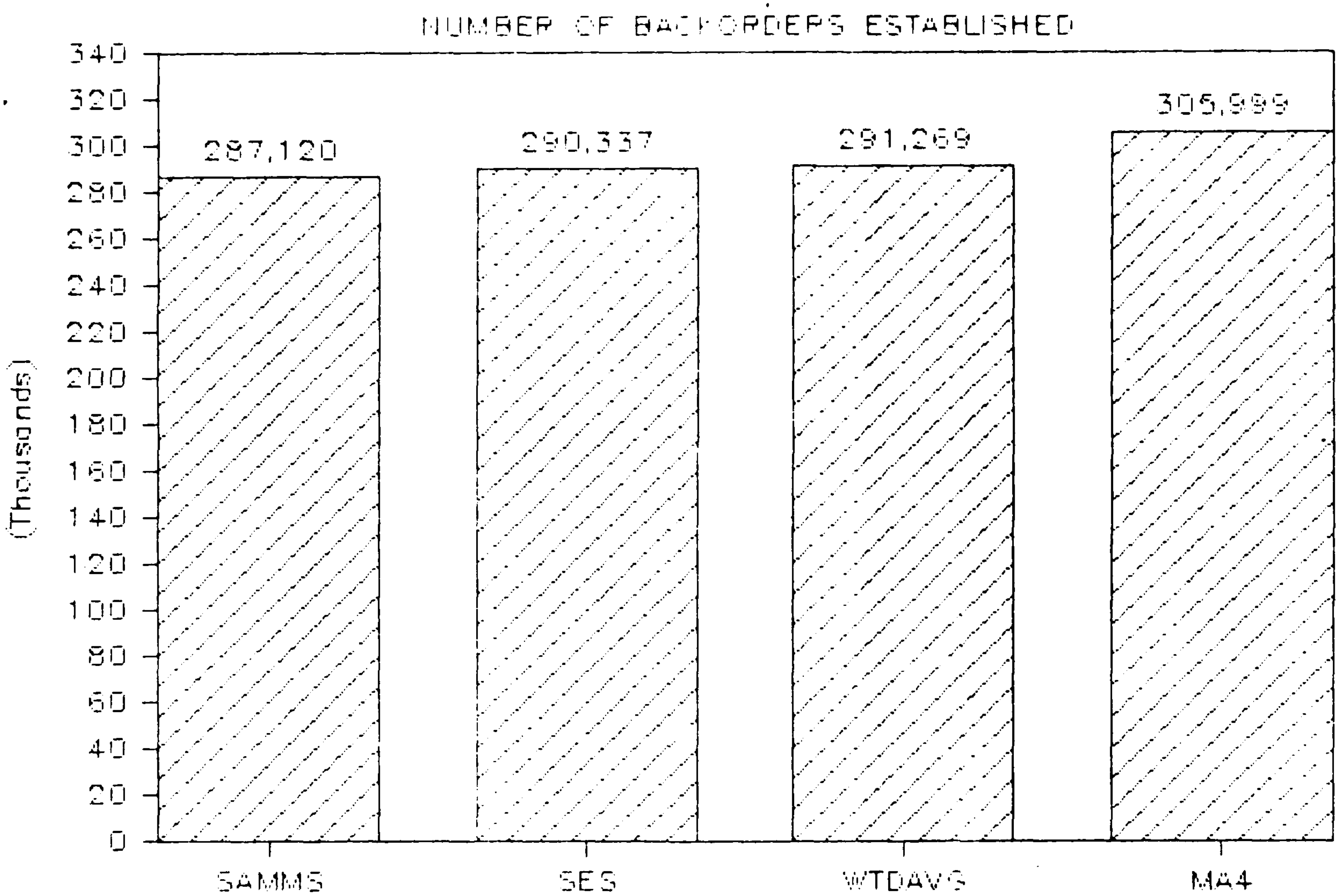
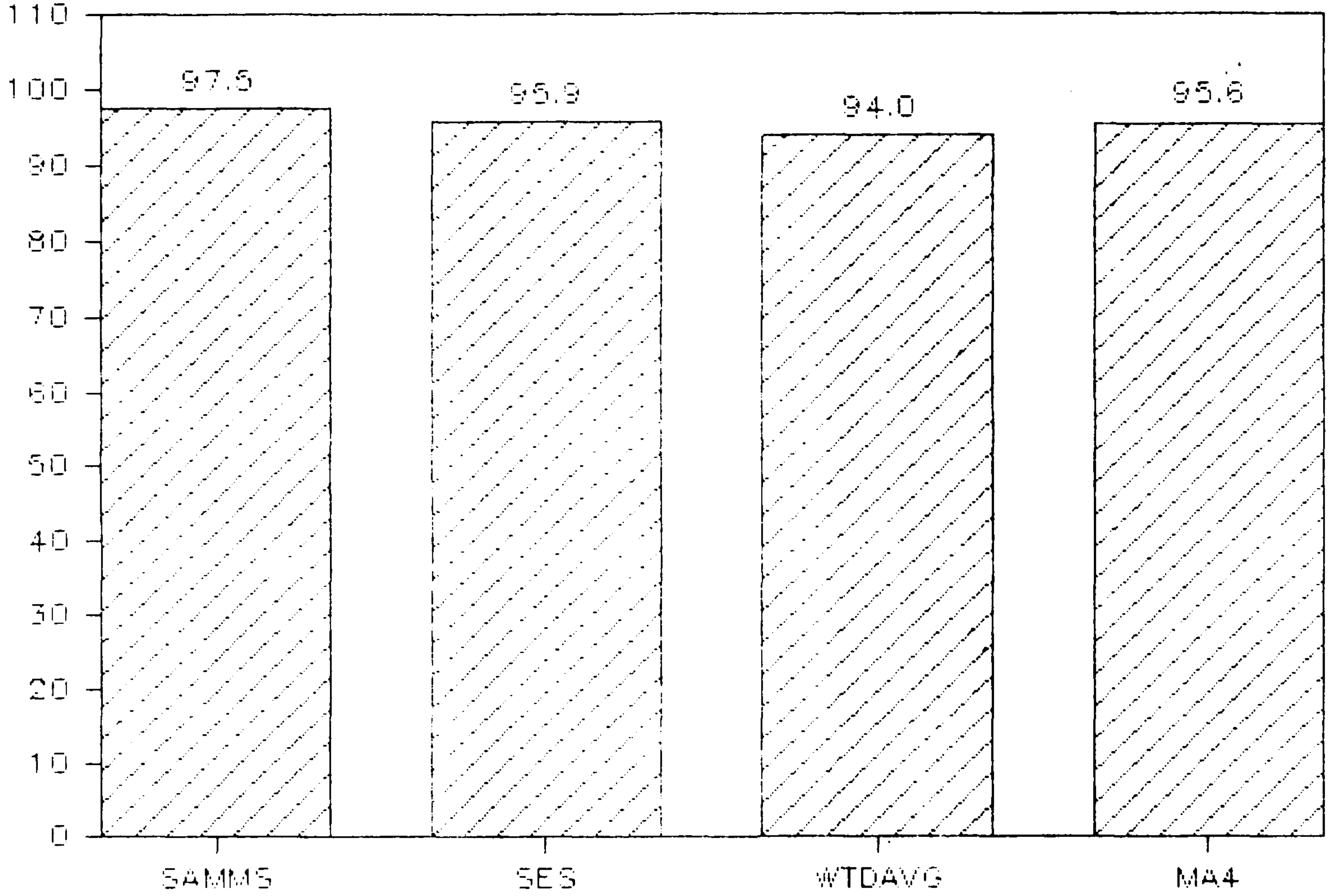

NUMBER OF BACKORDERS ESTABLISHED
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
(Thousands)
287,120
290,337
291,269
305,999
SAMMS
SES
WTDAVG
MA4

FIGURE 5

AVERAGE DAYS TO RELEASE BACKORDER



availability and days on backorder. It is these totals and averages which are presented in the figures.

Figure 1 shows the average supply availability over the eight quarters of the simulation run. As the figure shows, there is virtually no change in supply availability across the four methods. As noted previously, this is not necessarily a negative finding. If other measures can be improved with no decrease in availability, this is indeed a positive impact.

Figure 2 shows the average safety level dollars totaled over all commodities. Each of the three alternative methods resulted in lower safety level dollars than the baseline SAMMS method. The percentage decreases depicted in Figure 2 were 1.8%, 13.4%, and 18.8% for SES, WTDAVG, and MA4, respectively.

Figure 3 presents the average value of commitments (in dollars). Once again, the three alternative methods produce smaller totals than the SAMMS method. The percentage decreases for SES, WTDAVG, and MA4 were 32.1%, 24.4%, and 14.5%, respectively.

Figure 4 shows the average number of backorders for each quarter. Here, all three alternative methods produced more backorders than the SAMMS method. The percentage increases were 1.1%, 1.4%, and 6.6% for SES, WTDAVG, and MA4, respectively.

Finally, Figure 5 depicts the average number of days to release a backorder. The SAMMS baseline method was again the poorest performer. The percentage decrease in average days to release a backorder was 1.6% for SES, 3.6% for WTDAVG, and 2.0% for MA4.

Taking these results as a whole shows that the weighted average appears to be the most consistent of the three alternative methods. While SES did well on commitments, it did poorly on safety level dollars. Similarly, the MA4 method performed well on safety level dollars, but poorly on commitments and backorders. If all of these variables are considered equal in importance, the WTDAVG appears to be the best choice of the alternatives examined. With the exception of number of backorders, this method was clearly superior to the current SAMMS forecasting procedure.

F. Results of Supplemental Analyses

In addition to the analyses and findings presented to this point, there were several additional issues which were considered in the study. Data were analyzed separately for each of these issues; the results of these supplemental analyses are presented in this section.

1. Monthly vs. Quarterly Forecasting

The first of these issues examined was whether monthly forecasting of selected (VIP) items was beneficial. To address this issue, the USIMS model was used to assess the impacts of eliminating monthly forecasting of VIP items. The analysis was accomplished by comparing the current procedure of monthly forecasts for VIP items to two alternatives: forecasting all items quarterly, and forecasting all items monthly.

Table 14 shows the impacts of switching to quarterly forecasting for all items and monthly forecasting for all items for the five variables discussed previously. The second and third columns of the table show that there is virtually no change in supply availability resulting from the switch to quarterly forecasts. The number of backorders and the days to release a backorder increase slightly (less than 1%) when all items are forecasted quarterly. The largest changes are the decreases in safety level dollars (5.3%) and average commitments (12.2%).

The last two columns of Table 14 show the impacts of switching to monthly forecasting for all items. Essentially, the impacts are just the opposite of those observed for quarterly forecasting of

Table 14

EFFECT OF FORECASTING
ITEMS ON A QUARTERLY MONTHLY BASIS

	CURRENT SYSTEM	ALL ITEMS QUARTERLY	PERCENT CHANGE	ALL ITEMS MONTHLY	PERCENT CHANGE
NUMBER OF BACKORDERS	287119.90	289132.90	0.70	282957.40	-1.48
SAFETY LEVEL DOLLARS	1661277.20	1572954.20	-5.26	1747506.10	5.26
SUPPLY AVAILABILITY	94.96	94.82	-0.15	94.98	0.02
DAYS TO RELEASE BACKORDER	97.53	97.94	0.42	91.92	[illegible]
TOTAL AVERAGE COMMITMENTS	2094222.60	1839179.00	[illegible]	[illegible]	[illegible]

all items. That is, the number of backorders and days on backorder both decrease, while safety level dollars and total commitments increase (the increase is rather dramatic in the case of total commitments). Again, supply availability remains virtually unchanged.

These results suggest a somewhat linear relationship between the proportion of items forecasted monthly and the inventory variables considered. Specifically, the more items subjected to monthly forecasts, the lower backorders and days on backorder will be, but the greater safety levels and commitments will be. The figures shown in Table 14 seem to indicate that the magnitude of the reductions in safety level and commitments associated with quarterly forecasting of all items more than offset the slight corresponding increases in backorders and days on backorder.

2. Effects of Including Foreign Military Sales in Forecasts

The next issue examined was whether or not to include foreign military sales (FMS) in the forecast. FMS are currently excluded from the forecasts.

The SAMMS version of double exponential smoothing was used to analyze the impacts of adding FMS to the forecasts. Individual values were obtained for the smoothing term, alpha, and these values were used in the comparisons. The backcasting technique was utilized to start each of the forecasts. One-step ahead forecast errors were calculated for all periods of demand available for each item. The absolute values of these were then summed and divided by the number of periods to create the measure of error, the Mean Absolute Deviation of Forecast Errors (MAD).

The items used in the analysis were the 6,412 from Sample 1. The comparison of forecast accuracy when FMS are included and excluded was accomplished using nonparametric statistical tests. Nonparametric tests, which usually use ranks instead of raw scores, do not require assumptions about the form of the distribution of the data, and are therefore appropriate for use in the present context. The test used here was an equivalent of the Wilcoxon Signed Ranks Test for paired comparisons. The test was used to compare the average error when FMS are included or excluded from the forecast computation. This was accomplished by ranking the errors within each item, computing the difference between the ranks for each item, and comparing this difference score with zero. If one forecast error was not consistently higher than the other, the expected value of the difference score would be zero. A t-test was used to determine whether the observed difference score was statistically significant from zero.

The average forecast error produced using the procedures of the current SAMMS forecasting method was 158.94. When FMS demand was included, the average forecast error increased to 165.15. The difference between the two mean ranks (1.74 when FMS was included, 1.26 when it was not) was significantly different from zero ($t = 59.89$, $p < .001$), indicating that the inclusion of FMS produces a forecast with significantly greater error than the current SAMMS procedure of excluding such demand.

3. Effects of Excluding Nonrecurring Demand in Forecasts

The third and final question addressed in these analyses was whether changing the current procedures for handling nonrecurring demands would increase forecast accuracy. Currently, SAMMS uses a portion of the nonrecurring demand to forecast high demand value items, and all of the nonrecurring demand to forecast medium and low demand value items. The analysis here examined the alternatives of (a) including all nonrecurring demand for all items, and (b) using only recurring demand to forecast all items. The 6,412 items from Sample 1 were used for the analysis. Of these, 42 items had all nonrecurring, and no recurring, demand. These were dropped from the analysis, leaving 6,370 items.

For the nonrecurring demand analysis, Friedman's two-way analysis for blocked designs was employed. This nonparametric method uses an approach which is similar to analysis of variance, but with the resulting test statistic approximating a chi-square distribution, rather than the F distribution of the parametric ANOVA procedure. If the computed chi-square value is statistically significant, the usual procedure is to compare the means of the different groups to determine which means differ significantly from each other. For this analysis, the procedure used is equivalent to Fisher's Least Significant Difference (LSD) method, but using ranks rather than the raw data.

The results of the Friedman procedure showed a significant difference between the mean ranks of the three sets of forecast errors, $\chi^2 = 304.6$, $p < .001$. The mean ranks, along with the mean raw errors, are shown in Table 15 below.

Table 15

MADs FOR ALTERNATIVE
TREATMENTS OF NONRECURRING DEMAND

Method[a]	Average MAD	Average Rank
SAMMS	212.9	2.10
All Nonrecurring Demand	221.8	2.08
Recurring Demand Only	211.4	1.82

[a]See text for a description of the various methods.

Post-hoc tests of the rank scores showed that using recurring demand only produced a significantly lower average rank error than either alternative. The average error scores, however, provide a slightly different picture. As the table shows, using all nonrecurring demand for all items produces the largest average error. In addition, using only recurring demand for all items, rather than the current SAMMS procedure, results in a slightly lower average forecast error (the difference between the two means is less than 1%).

One possible explanation for the above finding is that nonrecurring demand is more erratic, and therefore more difficult to forecast, than recurring demand. To examine this hypothesis, those items which had extreme reductions in forecast error when comparing the current system to using recurring demand only were identified. Plots of the forecasts and actual demands using both demand streams were then examined. As expected, the demand plots when nonrecurring demand was included were much more erratic than those excluding nonrecurring demand. As a result, the forecasts based on recurring demand only were slightly more accurate than those based on both types of demand.

To summarize, this section presented the results of analyses designed to determine the effects of (1) quarterly versus monthly forecasting, (2) including foreign military sales in the demand forecasts, and (3) changing the way nonrecurring demand is treated in the forecasts. The results suggest positive benefits associated with changing current SAMMS procedures to forecast all items on a quarterly basis. There is also some indication that using recurring demand only to forecast all demand may produce a lower forecast error than the current procedure of incorporating nonrecurring demand in the forecasts.

VI. DISCUSSION AND SUMMARY

The results presented here represent a number of different analyses covering a wide range of areas. This section will attempt to summarize the key findings of the study, to present explanations and interpretations for these findings, and to suggest areas for further research.

A. Results Of Statistical Analyses

Prior to a discussion of the results, it is necessary to consider briefly the nature of the data itself. As noted in the previous section, the analysis of the ACFs for the items in the three samples indicated that at least 70% of the items to be forecasted had demand patterns which were random. Although the ACF analysis carried out here is admittedly a weak indicator of the existence of patterns the analysis does illustrate an important factor regarding the data to be forecasted. For the majority of items, the historical demand data is not a reliable basis upon which to forecast future demand.

The above conclusion does, of course, have serious implications for the choice of a single technique to be used to forecast all DLA items. The current SAMMS version of double exponential smoothing, along with all of the alternative forecasting techniques tested in this study, utilize past data to predict the future. If the past demand history for an item is not representative of the future, then the forecast accuracy of any of these methods will obviously be compromised. This issue should be considered in any interpretations of the findings of the present study.

A total of 18 forecasting methods were actually compared in the study. These methods were selected based on an examination of a much larger number of techniques, representing a wide range of forecasting approaches. A thorough literature search identified models used in the past by DLA and by the Services, along with methods which have shown promise in the academic literature. It is believed that this procedure represented a comprehensive assessment of the current knowledge regarding forecasting in inventory environments, and included recent trends (such as the use of averaging of forecasts) in the literature as well. Some of the techniques examined were rejected as being too costly and complex to be practical in the large inventory environment. The 18 methods finally selected were those which showed the most promise in the literature, and were best suited to the needs of DLA as well.

The results of the analyses of all 18 of these methods were reported in Tables 3-5. Several findings reported in these tables are notable. First, the use of decomposition in order to eliminate seasonality from the data did not result in very accurate forecasting. This is not surprising, given that the autocorrelation analyses alluded to earlier showed that relatively few items (1.5% of Sample 1) appeared to have seasonal demand patterns. It is, in fact, difficult to imagine very many items, outside of some subsistence and clothing items (not included in this study), which might be seasonal in nature. Thus deseasonalizing the data for all items is not a very effective forecasting procedure.

Methods which are designed to handle trends in the data also tended to be poor performers. Again, only about 2% of the items in the first sample were judged to have trends in their demand streams. This explains the relatively poor performance of the double exponential smoothing methods (Brown's and Holt's), and the superior performance of Gardner's model (which damps the trend term) over Holt's.

As would be expected, the more accurate forecasting methods were those which ignore both trend and seasonality in the data. These methods include single exponential smoothing, the 4- and 8-quarter moving averages, and the current SAMMS method (Brown's double exponential smoothing without the term that adjusts for the trend).

As a result of these findings, the 18 methods were narrowed down to six: single exponential smoothing, 4-quarter moving average, naive, mean, Holt's exponential smoothing, and the naive using deseasonalized data. The choice of these six methods represented a compromise between several considerations: which methods perform best overall (single exponential smoothing and the moving average), which method is the single best method for the largest number of items (naive), and which subset of methods when considered together complement each other so as to produce the smallest error (the remaining four). Each of these represents a different perspective on the issue of how to determine the adequacy of a forecasting method.

The issue of how to determine forecast accuracy illustrates another finding of the study. Clearly, different forecasting methods are more successful with different items, and overall forecast accuracy could be improved by using multiple forecast methods, and matching methods with particular items. One problem here is that using many different methods is very costly; it would be preferable to identify a few methods which would accurately predict most items. The problem then becomes how to match items and forecast methods, and how to predict, for a new item, which forecast method will work best.

There are several approaches which could be taken in solving this problem. The procedure used here was to form items into groups based on the similarity of the forecasting method which produced the smallest error. Once the groups are established, item characteristics must be identified which allow for the classification of new items into one of the groups.

The method used in this study to accomplish the above-described task was a multivariate statistical procedure known as discriminant analysis. This procedure is ideally suited to the problem, since it allows for the selection of a small group of characteristics from a larger pool, and also provides a way of using these characteristics to classify items.

The results of the discriminant analysis procedure showed that overall, the prediction of forecast groups based on item characteristics was quite poor. There are at least two possible explanations for this finding. First, the item groupings, which were formed based on the forecast error, may not have been meaningful. If the items within the groups were not in fact homogeneous with regard to the variables used to predict the groups, then prediction would be expected to be poor. In other words, if the items within a particular group were no more similar to each other than they were to items in other groups, prediction of item groupings would be difficult.

The other explanation for the poor prediction relates to the item characteristics themselves. It may be that the item groupings are meaningful, but that none of the characteristics chosen are good predictors of these groupings. This implies that there may still be some variable or set of variables which could be used to successfully predict the item groupings.

Of these two explanations, the former is probably the more reasonable in this instance. Given the great degree of variability in the data, it appears from the results that grouping items by forecast method does not result in homogeneous groups which can then be predicted by other variables. This conclusion in turn suggests the alternative procedure of forming item groupings based on the characteristics of the items themselves. Once the groups were formed, the "best" forecasting method for each group could be identified. There is, however, no reason to suspect that this approach would have been any more successful than the one used here.

The data presented in Table 7 do lend some support for the idea of using multiple forecasting methods rather than a single method for all items. If the prediction of which methods to use with which items could be made with a reasonable degree of accuracy, forecast error could be reduced. The study's findings showed that the more forecasting methods that were used, the lower the forecast error, given perfect prediction. However, as the number of forecast methods/item groupings increased, the ability to discriminate between the groups decreased, as did forecast accuracy.

If the prediction of item groupings could be improved, the forecasting method developed as a part of this study would also be more effective. This method used the discriminant analysis procedure to select one or the other of the forecasting methods used in the average when the choice can be made with a reasonable degree of confidence. If the prediction of the groups could be improved, this method might prove to be a more effective alternative to simply averaging forecasts from multiple methods. Ultimately, the statistical analysis was carried out on the entire population of forecasted items, using a small subset of the methods originally considered. These results showed that the weighted average of the forecasts from single exponential smoothing and the four quarter moving average produced the greatest improvement over the current SAMMS forecasting method.

The finding that the weighted average is the best performer overall for the population of items is consistent with the recent forecasting literature. Makridakis and Winkler (22), for example, note that lacking a theoretical or other strong basis for choosing a particular forecasting method, averaging several methods may produce a superior forecast. Since there is no compelling reason for choosing one method over another here, and since the efforts to match forecast methods to items were not too successful, averaging represents the next logical choice for obtaining improved forecast accuracy.

In terms of implementation, the use of the weighted average of two forecasts has both advantages and disadvantages. On the positive side, use of the average means that several distinct techniques can be examined on a continuing basis. That is, forecasts using SES and MA4 could be used individually, an unweighted average could be examined, and the discriminant analysis method developed here could also be tested. Since demand patterns appear to be highly unstable, it is possible that the weighted average technique might not be the most effective of these at some point in the future.

One possible disadvantage of the WTDAVG technique is the increased processing time and storage space required. Two forecasts must now be calculated, although the calculations are no more involved than those for the single and double smoothed values already computed as part of the current SAMMS procedure. These must then be combined based on the relative magnitudes of the forecast errors associated with the two methods. This requires storing two forecasts rather than one, in addition to storing the last four quarters of demand required for the moving average. Given the current power of computer hardware to store and process information, however, the impacts of the extra requirements associated with this method would be not be significant.

B. Findings Of Analysis Of Inventory System

In order to assess the impacts of the forecasting methods on inventory system variables, the study made use of a simulation model (USIMS). Figures 1-5 presented the results of these analyses in terms of several key variables associated with the inventory system.

In the current context, the most useful conclusion to be drawn from the simulation results is that they lend additional support for the superiority of the weighted average technique. This method, when compared with the current SAMMS technique, resulted in lower safety level dollars and total commitments while maintaining the same level of supply availability. By increasing forecast accuracy, it is no longer necessary to maintain the same amount of safety stock, which provides a shield against those errors. At the same time, greater accuracy can translate into lower commitments. As overforecasting (a more common problem than underforecasting) is reduced by improving accuracy, the amount of stock purchased will also decrease, representing a one-time savings in commitments.

The only variable for which the WTDAVG resulted in poorer performance was number of backorders. There are at least two possible explanations for this finding. First, there were several parameters in the simulation model which were held constant and which affect the number of backorders. The system constant, reflecting the dollar value of the MAD, was not changed from one method to the next. The backorder goal (beta), which interacts with the system constant to influence the number of backorders, was also held constant. Thus there are other parameters in the SAMMS system which affect the inventory levels, but whose impact was not directly examined as part of the simulation analysis.

Another possible explanation for the increase in backorders relates to the observed decrease in the safety level. The problem relating to backorders is not so much inaccurate forecasts as it is demand variance. That is, some items demonstrate demand patterns that fluctuate wildly from one period to the next. Increasing overall forecast accuracy may be possible for such items, but the demand variance problem remains. The situation is made even worse by the decrease in safety level associated with the greater forecast accuracy. Now, the protection against demand variance has been reduced, thereby increasing the possibility for backorders. A comparison of Figures 2 and 4 shows that the greater the decrease in safety level from one method to the next, the greater the increase in number of backorders. Thus it would appear that reducing forecast error alone is not sufficient to reduce the number of backorders. To accomplish the latter, the problem of demand variance must be addressed (variance in leadtimes is another problem which should be examined, as it too may explain the results discussed above).

Several inconsistencies were observed in the results produced by the USIMS model. The most obvious of these is the difference in the performance of single exponential smoothing in the model versus the statistical analysis. SES appeared to do quite well in the simulation runs, although it did quite poorly in the statistical analysis. In the latter analysis, SES produced the largest MAD and RMSE of the eight forecasting methods compared for the population of items (see Table 10).

Within the USIMS model itself, another inconsistency is the magnitude of the supply availability figures, as shown in Figure 1. These appear to be unrealistically high, and this is an acknowledged problem with the model. It should be noted, however, that it is the relative value of this variable across methods that is of interest in the current study.

Yet a third inconsistency related to the change in the levels over the length of the simulation. An examination of the data by quarter was expected to reveal increasing differences between the methods, as the start-up effects wore off with time. This was not observed to be the case. No consistent pattern emerged in the levels across time, although the limited 8-quarter time horizon examined here may have been insufficient for the identification of such differences.

The inconsistencies noted above serve to underscore the preliminary nature of this analysis of the impacts of the forecasting methods on the supply system. The time constraints of the present effort required the use of an inventory model which had already been developed. The USIMS model was the best one available for the analysis. The model is, as noted previously, an imperfect duplication of the SAMMS system. The inconsistencies noted above suggest caution in drawing conclusions about the anticipated magnitude of the changes in inventory system levels which would accompany a change in forecast methods.

The most important insight that the USIMS simulation model did provide, which could not have been obtained from the statistical analysis alone, relates to the issue of implementation. In running the simulation analysis, various implementation issues needed to be addressed simply to produce the output. In addition, analysis of the USIMS results suggested several issues and questions concerning how best to implement a new forecasting technique.

One issue which needed to be addressed in order to actually run the simulation analysis was how to implement the new forecasting method. Some methods, including the weighted average, require starting levels. There are various techniques which could be used to start the new forecasting method. In the case of single exponential smoothing, for example, one could either begin with the current SAMMS single smoothed value (as was done in the simulation analysis), or backcast (as was done in the statistical analysis), or use an average of the last few quarters to determine a starting point. All of these methods would lead to different impacts on the system, and these impacts would be felt over varying lengths of time.

Another implementation issue relates to the fact that the effects of instituting a new method will not be felt immediately, but rather will be extended over time. Many items have long leadtimes, so that current system activity will be based on decisions made under the old forecasting method. Similarly, some items will have large amounts of excess stock which were acquired under the old forecasting system. In these cases, it will take a long time for reorder points to be reached and the benefits of a reduced forecast error to be realized. Thus the two years over which the USIMS analysis was run was probably not enough time to observe the total impacts on the inventory system.

Another consideration is whether or not to change, at least on a one-time basis, any other SAMMS calculations as part of the implementation of the new forecast method. Although the forecast error is an important determinant of the levels of key inventory system measures, there are many other factors which also enter into play.

The variable safety level, for example, is influenced not only by the MAD, but also by the item's leadtime, price, QFD (through the economic order quantity, EOQ), and average requisition size, along with the value of the system constant (dollar value of the MAD leadtime, MADLT, totaled for all items in each commodity). To the extent that the new forecasting technique results in changes in these other factors, the benefits of reducing the MAD may be enhanced or reduced.

Two factors which are affected by a change in forecast method and which might exert opposite influences on the safety level are the system constant and the QFD. A decrease in the MAD will produce a corresponding decrease in the system constant, thereby further reducing the safety level beyond the reduction associated with the MAD itself (this effect which was not accounted for in the simulation analysis presented here, since the same system constant was used for all methods). By contrast, a decrease in the QFD will produce a decrease in the EOQ, which in turn will increase the safety level. Thus it is conceivable that implementing a more accurate forecasting technique which produced a smaller forecast of demand could result in an increase in safety level for some items.

The relationship between reducing forecast error and improving other system measures is further complicated by the fact that the MAD is smoothed before it is converted to the MADLT value. That is, the new forecast error (MAD) is multiplied by the alpha value (usually .2), and this result is added to the previous MAD value. This smoothing obviously postpones any benefits to be gained by reducing the size of the forecast error. For purposes of implementation, therefore, the benefits of the new forecasting method might be best realized if the smoothing constant (alpha) was increased for the first few time periods. This does, however, need to be tested.

There are undoubtedly many other considerations relating to the implementation of a new forecasting technique into SAMMS. The issues discussed above are those which were apparent from a review of the results of the simulation analysis using the USIMS model. Clearly, these results raised more questions about these implementation issues than they answered. It does appear, however, that the way in which the forecasting method is integrated into the current SAMMS system is a key element in determining the extent of its impacts.

Given the importance of these implementation issues, it seems prudent to suggest that additional study should be given to these concerns prior to any implementation of the weighted average method (or any other alternative technique). The purpose of this study would be to compare alternative procedures for implementing the new forecasting method, comparing the short- and long-term impacts on inventory levels.

Another goal of this suggested implementation study would be to determine whether the new method should be implemented for selected commodities, selected items, or "across the board". This is an important consideration which is based on the results of the statistical analyses presented here. Any new forecasting method will not perform better than the old method for all items. When the impact on inventory variables such as safety level dollars and commitments is the measure of interest, it becomes important for the new method to perform better on the "right" types of items (for example, items with high annual dollar demand). An analysis of items on an individual basis is required to produce this detailed level of information. Again, this is an implementation issue which would best be resolved by a study specifically designed to examine these concerns.

Finally, this type of study could also be used to determine how best to implement a change from monthly to quarterly forecasting of VIP items. The results shown in Table 14 seem to suggest benefits, in terms of lower average commitments and safety level dollars, associated with a switch from monthly to quarterly forecasting of these items. Further study would help to determine which VIP items would benefit the most from quarterly forecasting.

C. Suggestions For Future Research

The findings of this study and the conclusions based on these results all point to the idea that "improving forecasting in DLA" is too vague and nonspecific a goal for future forecasting studies. The results show clearly that some methods will outperform others for some items; therefore, improvement in forecast accuracy can almost always be obtained for at least some items. By focusing in on a small group of items, it should be possible to demonstate some improvements in forecast accuracy. The important point here is that the items should be selected a priori, based on criteria which are consistent with the stated goals of increasing forecast accuracy. As an example, one goal of improved forecasting might be to reduce costs by decreasing safety level dollars. If this were the case, one strategy might be to focus in on high demand/high dollar value items, say the top 1% in each commodity. Another strategy might be to pick items which are already in "long supply", and attempt to determine the degree and manner by which current forecasting procedures created this situation, and alternative methods which could be used in the future.

The particular strategy selected is not crucial for purposes of future research. What is essential is that the focus of the study be narrowed, and that supply experts in DLA do the narrowing. The supply experts' interpretations of the agency's policy decisions must determine the scope of any future studies. The researcher's choice of items, variables, procedures, and error measures are all strongly influenced by the direction set forth by policy-makers in supply operations.

The type of approach discussed above would have several advantages. First, it would greatly increase the chances of discovering similarities among item characteristics, including demand patterns. In addition, it would concentrate research efforts in an area which is believed, by supply experts, to have the most potential benefit to the agency. Finally, narrowing the scope of the study allows results to be obtained more quickly. This in turn permits re-direction of efforts if areas currently being pursued do not appear to be fruitful.

Similarly, it should be possible to obtain a priori groups of items, rather than attempting to group items based on forecast method. It may be possible, for example, to identify items which should be seasonal, based on the items' function. This information could be used in conjunction with an analysis of the item's demand history to identify a group of seasonal items. Once this was accomplished, forecasting methods which are specifically designed to handle seasonality could be compared for only these items. In addition, item characteristics which might be used to predict an item's being seasonal could be explored.

Another hypothetical grouping might be items which all belong to the same weapon system. For these items, demand might be more successfully forecasted using program data, such as number of flying hours, and a regression analysis procedure.

This type of approach has the clear advantage that the item groupings are known to be meaningful ones, since they are based on a priori information. On the negative side, such an approach would be extremely time-consuming, and would probably prove to be useful for only a minority of DLA's items. Finally, there is again no guarantee that proceeding along these lines would produce any greater success than the methods used in the current study.

One final approach might also be useful to explore in future studies. Smith's focus forecasting (21) was initially considered for inclusion in the study, but was rejected due to lack of empirical evidence. The approach, however, does have a great deal of intuitive appeal, especially given the erratic and variable nature of the demand patterns for the items to be forecasted by DLA.

Focus forecasting is another technique which involves the use of multiple forecasting methods. Rather than averaging, focus forecasting uses the method that worked best for an item in the past to forecast that item in the future. The key ingredient of the focus forecasting approach is that the various methods used are very simple and intuitively obvious (these are the methods usually referred to as "naive" in the forecasting literature). Given the nature of the data, it might well be that such "naive" approaches would be the most successful. The methods which did work best in the present study tended to be the ones which were simpler.

The focus forecasting approach, as described by Smith, has an additional advantage in that it allows for input from the item managers. In the current SAMMS system, item managers can have a fairly large degree of influence on the forecasting process, if they choose to do so. Focus forecasting would allow the informal, but effective, methods used by item managers to be formalized into the forecasting system. This may not only improve forecast accuracy, but would have the added advantage of "de-mystifying" the forecasting process for the main users of the results of that process.

VII. CONCLUSIONS

The following conclusions are based on the results of the demand forecasting study documented in this report.

- Several simple forecasting techniques produce a lower forecast error than the current SAMMS method.

The results of the statistical analyses showed that single exponential smoothing, the four-quarter moving average, and a weighted average of the forecasts of these two methods all produced lower root mean squared errors and lower mean absolute deviation of errors than the current SAMMS method. This was true for all commodities except for the Medical commodity, where only the MA4 method produced a slight improvement over the current SAMMS procedure.

- A weighted average of the forecasts generated by single exponential smoothing and the four-quarter moving average produces the greatest improvement in forecast accuracy.

Based on an analysis which included all of the items in the population which met the forecast criteria, this weighted average method produced a 2.9% reduction in root mean squared error, and a 3.9% reduction in mean absolute deviation of forecast error. The weights for the averaging are based on the previous period's forecast errors of each method. The weighted average method did not improve the forecast error for items in the Medical commodity. For this commodity, the four-quarter moving average alone was the best method.

- Substitution of the weighted average procedure for the current SAMMS procedure would result in: no change in supply availability, reductions in safety level dollars, commitments, and days on backorder, and an increase in the number of backorders.

The results of the simulation analysis suggest the impacts described above. The analysis was considered preliminary, however, so that no firm conclusions can be drawn regarding the magnitude of these changes. The results do suggest that the decreases in levels would be large enough, and the increase in backorders small enough, to recommend the weighted average method over the current SAMMS procedure. Based on a decrease in safety level dollars proportional to a decrease in forecast error, improvement of the best method over the current method is estimated to be as follows:

Commodity	Percent Reduced Error	Estimated Reduced Safety Level ($)
Construction	1.1%	$ 2,715,950
Electronics	1.6	1,912,796
General	6.0	12,714,489
Industrial	4.7	5,885,924
Medical	3.7	607,486
C & T	1.4	1,586,742

- Further study is needed in order to determine the best strategies for implementation of any new forecasting method.

The simulation results raised several issues regarding how to best implement a new forecasting technique. A more detailed examination of these issues would be needed in order to ensure that the maximum benefit is obtained from increasing forecast accuracy.

- Quarterly forecasting of all items, including VIP items, would result in: no change in supply availability, reductions in safety level dollars and commitments, and slight increases in the number of backorders and the average days on backorder.

These findings were also based on the results of the simulation analysis. The reductions in levels appear to be large enough to justify switching to forecasting all items quarterly, despite the possibility of slight increases in the number of backorders.

- The inclusion or exclusion of nonrecurring demand from the SAMMS calculations has little impact on forecast error.

The current method used by SAMMS includes a portion of nonrecurring demand for high demand value items, and all nonrecurring demand for other items. The results of an analysis of changing this procedure showed that the mean absolute deviation of forecast errors was reduced only slightly (less than 1%) if recurring demand only was used in forecasting all items.

- Including foreign military sales in the forecast of demand would result in greater forecast error.

The current SAMMS calculations exclude foreign military sales from consideration in forecasting demand. The results of an analysis of changing this procedure showed that the average forecast error would be increased if this type of demand were included in the forecast.

VIII. RECOMMENDATIONS

- The weighted average of the forecasts from single exponential smoothing and the four-quarter moving average methods should be implemented as the SAMMS forecasting system for all commodities except Medical.

This recommendation applies only to those items studied. This excludes new items, Clothing and Textile's Program Oriented Items and government furnished materiel, and subsistence items. It is anticipated that implementation of this method would reduce costs to DLA in the form of commitments and safety level dollars.

Customer service, as measured by supply availability, would not be adversely affected.

Since this new method did not improve forecast accuracy for items in the Medical commodity, it cannot be recommended for these items. For this commodity, the alternatives are to keep the current SAMMS forecasting technique, or to implement the four-quarter moving average, which produced a slight reduction in the average forecast error for this commodity.

- Further study of implementation issues should be undertaken prior to the incorporation of this or any other alternative forecasting method into the SAMMS system.

The goal of such a study would be to determine how best to implement the new method so as to maximize the benefits of improving forecast accuracy.

- Forecasting of all items should be carried out on a quarterly basis.

Eliminating the monthly forecasting of VIP items would result in decreased safety level dollars and commitments, and have no effect on supply availability.

Appendix A

Formulae for Forecasting Methods

This appendix provides the formulae for the forecasting methods examined in this study. In the formulae which follow, X_t is the actual demand for time period t, and F_{t+1} is the forecast for time period t+1.

1. <u>Naive forecast</u>

$$F_{t+1} = X_t$$

2. <u>Simple mean of past observations</u>

$$F_{t+1} = \sum_{i=1}^{t} X_i \,/\, t$$

3. <u>N-period moving average</u>

$$F_{t+1} = \sum_{i=k}^{t} X_i \,/\, n \,,$$

where n = number of periods in the moving average
k = t-(n-1).

4. <u>Single Exponential Smoothing</u>

$$F_{t+1} = \alpha X_t + (1-\alpha)F_t \,,$$

where α is the smoothing constant (usually between Ø and 1).

5. <u>Current DLA version of exponential smoothing</u>

$$S'_t = \alpha X_t + (1-\alpha)S'_{t-1} \,,$$

$$S''_t = \alpha S'_t + (1-\alpha)S''_{t-1} \,,$$

$$F_{t+1} = 2S'_t - S''_t \,,$$

where S'_t is the single smoothed value and S''_t is the double smoothed value.

6. <u>Brown's double exponential smoothing</u>

$$S'_t = \alpha X_t + (1-\alpha)S'_{t-1} \,,$$

$$S''_t = \alpha S'_t + (1-\alpha)S''_{t-1} \,,$$

$$a_t = 2S'_t - S''_t \,,$$

$$b_t = \alpha/1-\alpha (S'_t - S''_t),$$

$$F_{t+m} = a_t + b_t m \,,$$

where a_t is the estimate of the level of the series, b_t is the estimate of the trend, and m is the number of periods ahead to be forecast.

7. Holt's double exponential smoothing

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + b_{t-1}) ,$$

$$b_t = \beta (S_t - S_{t-1}) + (1-\beta) b_{t-1} ,$$

$$F_{t+m} = S_t + b_t m ,$$

where β is the smoothing constant for the trend term (b_t).

8. Gardner's double exponential smoothing

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + \varphi b_{t-1}) ,$$

$$b_t = \alpha (S_t - S_{t-1}) + (1-\alpha) \varphi b_{t-1} ,$$

$$F_{t+m} = S_t + \sum_{i=1}^{m} \varphi^i b_t ,$$

where φ is an additional smoothing constant used to "damp" the trend term.

9. Trigg-Leach adaptive exponential smoothing

$$F_{t+1} = \alpha_t X_t + (1-\alpha_t) F_t ,$$

where $\alpha_t = |E_t/M_t| ,$

$$E_t = \beta e_t + (1-\beta) E_{t-1} ,$$

$$M_t = \beta |e_t| + 1-\beta (M_{t-1}) ,$$

$$e_t = X_t - F_t ,$$

α_t and β are smoothing constants and $|\ |$ denotes absolute value.

Appendix B

Selected Characteristics

of Population and Samples

Table B-1

COMMODITIES FOR POPULATION AND SAMPLES

Commodity	Population	Percentage Sample 1	Sample 2	Sample 3
Construction	14.1	14.7	13.9	14.5
Electronics	25.7	25.4	25.4	24.3
General	13.0	12.4	13.2	12.7
Industrial	43.1	43.2	43.8	44.6
Medical	1.8	2.0	1.6	1.9
C & T	2.2	2.2	2.1	2.1

Table B-2

SUPPLY STATUS CODES (SSC) FOR POPULATION AND SAMPLES

SSC	Population	Percentage Sample 1	Sample 2	Sample 3
A	6.7	6.3	6.5	6.9
1	86.4	87.2	86.4	86.5
4	0.5	0.5	0.5	0.5
5	0.1	0.1	0.1	0.1
6	5.8	5.3	6.2	5.5
7	0.2	0.3	0.2	0.3
8	0.2	0.3	0.1	0.2

Table B-3

NUMBER OF QUARTERS OF DEMAND FOR POPULATION AND SAMPLES

Demand Qtrs	Population	Percentage Sample 1	Sample 2	Sample 3
32	52.2	52.3	52.3	52.9
28	11.8	11.8	11.8	11.4
24	3.3	3.4	3.4	3.1
20	2.8	3.2	3.2	3.1
16	14.4	13.7	13.7	14.3
12	12.2	12.6	12.6	12.0
8	2.2	2.2	2.2	2.3
4	1.0	0.9	0.9	1.0

Appendix C

Individual Parameter

Values for Smoothing Methods

SES ALPHA

P1	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0.1	4163	4163	64.925	64.925
0.2	708	4871	11.042	75.967
0.3	445	5316	6.940	82.907
0.4	338	5654	5.271	88.178
0.5	163	5817	2.542	90.721
0.6	110	5927	1.716	92.436
0.7	93	6020	1.450	93.886
0.8	73	6093	1.138	95.025
0.9	319	6412	4.975	100.000

DES ALPHA

P2	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0.1	5009	5009	78.119	78.119
0.2	675	5684	10.527	88.646
0.3	252	5936	3.930	92.576
0.4	138	6074	2.152	94.729
0.5	66	6140	1.029	95.758
0.6	29	6169	0.452	96.210
0.7	26	6195	0.405	96.616
0.8	28	6223	0.437	97.052
0.9	189	6412	2.948	100.000

TRIGG-LEACH BETA

P3	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0.1	2946	2946	45.945	45.945
0.2	480	3426	7.486	53.431
0.3	337	3763	5.256	58.687
0.4	278	4041	4.336	63.022
0.5	219	4260	3.415	66.438
0.6	163	4423	2.542	68.980
0.7	193	4616	3.010	71.990
0.8	196	4812	3.057	75.047
0.9	316	5128	4.928	79.975
1	1284	6412	20.025	100.000

SAMMS ALPHA

P4	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0.1	4646	4646	72.458	72.458
0.2	691	5337	10.777	83.235
0.3	265	5602	4.133	87.367
0.4	116	5718	1.809	89.177
0.5	79	5797	1.232	90.409
0.6	54	5851	0.842	91.251
0.7	44	5895	0.686	91.937
0.8	23	5918	0.359	92.296
0.9	494	6412	7.704	100.000

DECOMPOSED SES ALPHA

P5	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	86	86	1.341	1.341
0.1	472	558	7.361	8.702
0.2	59	617	0.920	9.623
0.3	31	648	0.483	10.106
0.4	21	669	0.328	10.434
0.5	22	691	0.343	10.777
0.6	13	704	0.203	10.979
0.7	12	716	0.187	11.167
0.8	11	727	0.172	11.338
0.9	5685	6412	88.662	100.000

DECOMPOSED DES ALPHA

P6	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	86	86	1.341	1.341
0.1	333	419	5.193	6.535
0.2	38	457	0.593	7.127
0.3	30	487	0.468	7.595
0.4	38	525	0.593	8.188
0.5	295	820	4.601	12.789
0.6	213	1033	3.322	16.110
0.7	181	1214	2.823	18.933
0.8	216	1430	3.369	22.302
0.9	4982	6412	77.698	100.000

DECOMPOSED SAMMS ALPHA

P7	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	86	86	1.341	1.341
0.1	593	679	9.248	10.590
0.2	60	739	0.936	11.525
0.3	41	780	0.639	12.165
0.4	39	819	0.608	12.773
0.5	38	857	0.593	13.366
0.6	43	900	0.671	14.036
0.7	43	943	0.671	14.707
0.8	38	981	0.593	15.299
0.9	5431	6412	84.701	100.000

DECOMPOSED TRIGG-LEACH BETA

P8	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	86	86	1.341	1.341
0.1	1725	1811	26.903	28.244
0.2	158	1969	2.464	30.708
0.3	78	2047	1.216	31.925
0.4	43	2090	0.671	32.595
0.5	19	2109	0.296	32.891
0.6	10	2119	0.156	33.047
0.7	8	2127	0.125	33.172
0.8	5	2132	0.078	33.250
0.9	1	2133	0.016	33.266
1	4279	6412	66.734	100.000

HOLT ALPHA

P9	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	420	420	6.550	6.550
0.1	2736	3156	42.670	49.220
0.2	1133	4289	17.670	66.890
0.3	606	4895	9.451	76.341
0.4	461	5356	7.190	83.531
0.5	289	5645	4.507	88.038
0.6	184	5829	2.870	90.908
0.7	132	5961	2.059	92.966
0.8	97	6058	1.513	94.479
0.9	62	6120	0.967	95.446
1	292	6412	4.554	100.000

HOLT GAMMA

P10	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1260	1260	19.651	19.651
0.1	2168	3428	33.812	53.462
0.2	623	4051	9.716	63.178
0.3	398	4449	6.207	69.386
0.4	280	4729	4.367	73.752
0.5	180	4909	2.807	76.560
0.6	179	5088	2.792	79.351
0.7	173	5261	2.698	82.049
0.8	185	5446	2.885	84.934
0.9	167	5613	2.604	87.539
1	799	6412	12.461	100.000

GARDNER ALPHA

P11	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	2796	2796	43.606	43.606
0.1	2342	5138	36.525	80.131
0.2	349	5487	5.443	85.574
0.3	221	5708	3.447	89.021
0.4	137	5845	2.137	91.157
0.5	118	5963	1.840	92.998
0.6	78	6041	1.216	94.214
0.7	73	6114	1.138	95.352
0.8	48	6162	0.749	96.101
0.9	40	6202	0.624	96.725
1	210	6412	3.275	100.000

GARDNER GAMMA

P12	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	758	758	11.822	11.822
0.1	367	1125	5.724	17.545
0.2	284	1409	4.429	21.974
0.3	198	1607	3.088	25.062
0.4	130	1737	2.027	27.090
0.5	198	1935	3.088	30.178
0.6	128	2063	1.996	32.174
0.7	172	2235	2.682	34.857
0.8	168	2403	2.620	37.477
0.9	304	2707	4.741	42.218
1	3705	6412	57.782	100.000

GARDNER PHI

P13	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1550	1550	24.173	24.173
0.1	378	1928	5.895	30.069
0.2	394	2322	6.145	36.213
0.3	419	2741	6.535	42.748
0.4	649	3390	10.122	52.870
0.5	358	3748	5.583	58.453
0.6	253	4001	3.946	62.399
0.7	217	4218	3.384	65.783
0.8	276	4494	4.304	70.087
0.9	704	5198	10.979	81.067
1	1214	6412	18.933	100.000

Appendix D

Example of One-Step and

Long-Term Forecast Computations

This study generated two different forecasts, a one-step ahead forecast and a long-term forecast, for time periods 29-32. Each of these forecasts is illustrated here, using both single exponential smoothing (SES) and Brown's double exponential smoothing (DES).

I. SES EXAMPLE

The formula for SES is:

$$F_{t+1} = \alpha X_t + (1 - \alpha) F_t,$$

where F_{t+1} is the forecast for the next period
X_t is the actual demand for the current period, t
α is the smoothing factor
F_t is the forecast for the current period.

Table D-1 presents some hypothetical demand data, along with the corresponding one-step ahead and long-term forecasts. It is assumed that the one-step forecast for t=28 was already calculated, and that α = 0.2.

Table D-1

SES EXAMPLE

t	X	F(one-step)	F(long-term)
28	100	97.0	-
29	105	97.6	97.6
30	107	99.1	97.6
31	114	100.7	97.6
32	120	103.4	97.6

The one-step ahead forecast of 97.6 for period 29 is obtained as follows:

$$F_{29} = \alpha X_{28} + (1 - \alpha) F_{28}$$

$$= (.2)(100) + (.8)(97.0)$$

$$= 97.6$$

In the calculation of the one-step ahead forecasts, each new observation is used as it becomes available. So, the one step-ahead forecast for period 30 is obtained by using the observed demand for period 29:

$$F_{30} = \alpha X_{29} + (1 - \alpha) F_{29}$$
$$= (.2)(105) + (.8)(97.6)$$
$$= 99.1$$

The one-step ahead forecasts for periods 31 and 32 are obtained in a similar fashion.

The long-term forecasts are calculated under the assumption that we are now at t=28 and must forecast the demand for the next four periods. Since SES has no trend or seasonal terms, the forecast for period 29 as the forecast for periods 30-32, as is shown in the last column of Table D-1.

II. DES EXAMPLE

The formulas for Brown's double exponential smoothing are:

$$S'_t = \alpha X_t + (1 - \alpha) S'_{t-1}$$

$$S''_t = \alpha S'_t = (1 - \alpha) S''_{t-1}$$

$$a_t = 2S'_t - S''_t$$

$$b_t = \alpha / 1 - \alpha (S'_t - S''_t)$$

$$F_{t+m} = a_t + b_t m$$

where S'_t is the single smoothed value,
S''_t is the double smoothed value,
a_t is the estimate of the level of the series
b_t is the trend term
F_{t+m} is the forecast for m periods ahead.

Table D-2 presents the same data as Table D-1, along with the one step-ahead and long-term forecasts using DES. Again, it is assumed that the values for S' and S" have already been calculated, and that $\alpha = 0.2$.

Table D-2

DES EXAMPLE

t	X	S'	S"	a	b	F(one-step)	F(long-term)
28	100	90.0	75.0	105.0	3.7	-	-
29	105	93.0	78.6	107.4	3.6	108.7	108.7
30	107	95.8	82.0	109.6	3.4	111.0	112.4
31	114	99.4	85.5	113.3	3.5	113.0	116.1
32	120	-	-	-	-	116.8	119.8

The one step-ahead forecasts use each subsequent actual demand, just as they did in the SES case. For example, the forecast for period 30 of 111.0 uses the actual demand for period 29 as follows:

$$S'_{29} = \alpha X_{29} + (1 - \alpha)S'_{28}$$
$$= .2(105) + (1 - .2)90.0$$
$$= 93.0$$

$$S''_{29} = \alpha S'_{29} = (1 - \alpha)S''_{28}$$
$$= .2(93.0) + (1 - .2)75.0$$
$$= 78.6$$

$$a_{29} = 2S'_{29} - S''_{29}$$
$$= 2(93.0) - 78.6$$
$$= 107.4$$

$$b_{29} = \alpha/1-\alpha(S'_{29} - S''_{29}),$$
$$= .2/(1 - .2)(93.0 - 78.6)$$
$$= 3.6$$

$$F_{30} = a_{29} + b_{29}m$$
$$= 107.4 + 3.6(1)$$
$$= 111.0$$

The forecasts for the remaining periods are calculated in a similar manner. Note that since these are one step-ahead forecasts, the term m in the forecast formula is 1.

The long-term forecasts are shown in the last column of Table D-2. These forecasts make use of the trend term b, using a different multiplier for each period ahead to be forecasted. That is, the m term in the formula is 1 for the period 29 forecast, 2 for the period 30 forecast, and so on. Since we are currently at period 28, the remaining values in the formulas are the ones for this period. For example, the forecast of 116.1 for period 31 is obtained as follows:

$$F_{31} = a_{28} + b_{28}m$$
$$= 105.0 + 3.7(3)$$
$$= 116.1$$

Appendix E

REFERENCES

1. Brown, R. G., Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Englewood Cliffs, N.J., 1962.

2. Gardner, E. S., "The Strange Case of the Lagging Forecasts," Interfaces, Vol. 14, 1984, pp. 47-50.

3. Defense Supply Agency, DSA Materiel Management System: Requirements Study, Defense Supply Agency, Alexandria, VA, July 1963.

4. Defense Supply Agency, Report of Simulation of Various Demand Forecasting Techniques, Defense Supply Agency, Alexandria, VA, June 1968.

5. Praggy, R. J., Comparison of Eight Demand Forecasting Models, Master's Thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, September 1981.

6. Bilikam, R., Interim Report on DESC Demand Forecasting Study, Defense Logistics Agency, Operations Research Office, Defense Electronics Supply Center, Columbus, OH, October 1984.

7. Sheehan, T. J., Subsistence Demand Forecasting Study, Operations Research and Economic Analysis Office, Defense Logistics Agency, Alexandria, VA, November 1984.

8. Orchowsky, S. J., Report on Analysis of the Program Oriented Item System for Forecasting Clothing Items, Operations Research and Economic Analysis Office, Defense Logistics Agency, Alexandria, VA, January 1985.

9. Fortney, W. G., M. I. Altschul, M. C. Banner, I. J. Bier, & J. C. Goodwin, Study of Demand Forecasting for Secondary Items: Volume II, Boeing Computer Services Company, Vienna, VA, October 1983.

10. Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, & R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," Journal of Forecasting, Vol. 1, 1982, pp. 111-153.

11. Armstrong, J. S., "Forecasting by Extrapolation: Conclusions from 25 Years of Research," Interfaces, Vol 14, 1984, pp. 52-66.

12. Mahmoud, E., "Accuracy in Forecasting: a Survey, Journal of Forecasting, Vol. 3, 1984, pp. 139-159.

13. Gardner, E. S., "Exponential Smoothing: The State of the Art," Journal of Forecasting, Vol. 4, 1985, pp. 1-28.

14. Gardner, E. S. and E. McKenzie, "Forecasting Trends in Time Series," Management Science, Vol. 31, October 1985, pp. 1237-1246.

15. Eilon, S. and J. Elmaleh, "Adaptive Limits in Inventory Control," Management Science, Vol. 16, 1970, pp. B533-B548.

16. Roberts, S. D. and R. Reed, "The Development of a Self-Adaptive Forecasting Technique," AIIE Transactions, Vol. 1, 1969.

17. Whybark, D. C., "A Comparison of Adaptive Forecasting Techniques," Logistics and Transportation Review, Vol. 9, 1973, pp. 13-26.

18. Trigg, D. W. and A. G. Leach, "Exponential Smoothing with an Adaptive Response Rate," Operations Research Quarterly, Vol. 18, 1967, pp. 53-59.

19. Box, G. E. P. and G. M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1976.

20. Steece, B. M. and S. D. Wood, "An ARIMA-Based Methodology for Forecasting in a Multi-Item Environment," in Forecasting: TIMS Studies in the Management Sciences, Vol. 12, S. Makridakis and S. C. Wheelwright, eds., North-Holland, New York, 1979.

21. Smith, B. T., Focus Forecasting: Computer Techniques for Inventory Control. CBI, Boston, 1978.

22. Makridakis, S. and R. L. Winkler, "Averages of Forecasts: Some Empirical Results," Management Science, Vol. 29, 1983, pp. 987-996.

23. Brandon, C. H. and C. M. Lackman, "Combined Forecast Based on Weighting Scheme Sharply Reduces Size of Error," Journal of Business Forecasting, Winter '84-'85, pp. 7-10, 16.

24. Granger, C. W. J. and R. Ramanathan, "Improved Methods of Combining Forecasts," Journal of Forecasting, Vol. 3, 1984, pp. 197-204.

25. Makridakis, S., S. C. Wheelwright, and V. E. McGee, Forecasting: Methods and Applications, John Wiley, New York, 2nd ed., 1983.

26. Armstrong, J. S., Long-Range Forecasting: From Crystal Ball to Computer, John Wiley, New York, 1978.

27. McNichols, C. W., An Introduction to Applied Multivariate Data Analysis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, 1980.

28. Cyrus, M. K., G. W. Arnett, S. E. Taylor, R. Parker, and E. L. Swim, Review of SAMMS Requirements Computations, Defense Logistics Agency, Operations Research and Economic Analysis Office, Cameron Station, Alexandria, VA, August 1985.

29. Naimon, S. G., Uniform SAMMS Inventory Management Simulation: Users' Guide, Defense Logistics Agency, Operations Research and Economic Analysis Office, Cameron Station, Alexandria, VA, January 1986.